

Regression on the Math Data: Part 1*

readmath2b.sas ends like this:

```
/* 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;
if ethnic=. then e6=.;
  else if ethnic=6 then e6=1;
  else e6=0;

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.'
      e6 = 'Other/DK 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;
label mtongue = 'English vs. Other';

/* Only use 2 of these if the model has an intercept! */
if course=. then c1=.; else if course=1 then c1=1; else c1=0;
if course=. then c2=.; else if course=2 then c2=1; else c2=0;
if course=. then c3=.; else if course=3 then c3=1; else c3=0;
label c1 = 'Catch-up' c2 = 'Mainstream' c3 = 'Elite';
```

* Copyright information is on the last page.

```

/* MathReg1.sas */
%include 'readmath2b.sas'; /* 2b has dummy variable definitions */
title2 'Variable Selection for Predicting Grade';

proc freq;
  title3 'Check dummy variables';
  tables sex*gender / norow nocol nopercent missing;
  tables tongue*mtongue / norow nocol nopercent missing;
  tables (e1-e4 e6) * ethnic / norow nocol nopercent missing;
  tables (c1-c3) * course / norow nocol nopercent missing;

proc reg;
  title3 'Model 1: Predict University Calculus Grade from HS Information';
  model grade = hsgpa hscalcalc hsengl;

/* It is very interesting to know what proportion of the remaining
  variation is explained by each variable, controlling for the other two.
  F = t-squared, and
  a = sF/(n-p + sF)
*/

proc iml;
  title3 'Proportion of remaining variation for HS information';
  n = 323; p = 4; s = 1;
  print "hsgpa controlling for hscalcalc and hsengl";
  t = 8.00; F = t**2; a = s*F/(n-p + s*F);
  print a;

  print "hscalcalc controlling for hsgpa and hsengl";
  t = 3.14; F = t**2; a = s*F/(n-p + s*F);
  print a;

  print "hsengl controlling for hsgpa and hscalcalc";
  t = -3.26; F = t**2; a = s*F/(n-p + s*F);
  print a;

proc reg;
  title3 'Model 2: Predict University Calculus Grade from Diagnostic Test';
  model grade = precalc calc;

proc reg;
  title3 'Model 3: Do the diagnostic test and HS info both contribute?';
  model grade = hsgpa hscalcalc hsengl precalc calc;
  Diagnostic_Test: test precalc=calc=0;
  HS_Information: test hsgpa=hscalcalc=hsengl=0;

proc iml;
  title3 'Proportion of remaining variation explained by diagnostic test';
  print "Precalc and calc controlling for hsgpa hscalcalc hsengl";
  n = 289; p = 6; s = 2; F = 8.28;
  a = s*F/(n-p + s*F); print a;

proc reg;
  title3 'Model 4: See if Course makes a contribution';
  model grade = hsgpa hscalcalc hsengl precalc calc c1 c3;
  Course: test c1=c3=0;
  Diagnostic_Test: test precalc=calc=0;

proc glm;
  title3 'Model 3 again with proc glm';
  class course;
  model grade = hsgpa hscalcalc hsengl precalc calc course;
  contrast 'Replicate Test of Course' course 1 -1 0,
  course 0 1 -1;
  contrast 'Diagnostic Test F = 9.06' precalc 1, calc 1;

```

```

options pagesize=100;
proc reg;
  title3 'Model 5: Include Language, Sex and Ethnic Background';
  model grade = hsgpa hscalcalc hsengl precalc calc
             mtongue gender e1-e4 e6;
  TroubleVars: test mtongue=gender=e1=e2=e3=e4=e6=0;
  Nationality: test e1=e2=e3=e4=e6=0;

proc reg;
  title3 'Model 6: Discarding Gender and Nationality';
  model grade = hsgpa hscalcalc hsengl precalc calc mtongue;
  EnglishTongue: test hsengl=mtongue=0;

proc iml;
  title3 'Proportion of remaining variation explained by mother tongue';
  print "Mtongue controlling for hsgpa hscalcalc hsengl precalc calc";
  n = 287; p = 7; s = 1; t = -2.23 ; F = t**2;
  a = s*F/(n-p + s*F); print a;

proc reg;
  title3 'Model 7: Drop mtongue and calc';
  title4 'Compare R-Square = 0.4556, Adj R-Sq = 0.4460 From Model 3';
  model grade = hsgpa hscalcalc hsengl precalc;

proc iml;
  title3 'Proportion of remaining variation explained by Pre-calculus';
  print "precalc controlling for hsgpa hscalcalc hsengl";
  n = 289; p = 5; s = 1; t = 3.63 ; F = t**2;
  a = s*F/(n-p + s*F); print a;

proc reg;
  title3 'Model 8: Combine precalc and calc instead of dropping calc';
  title4 'Compare R-Square = 0.4492 from Model 7';
  model grade = hsgpa hscalcalc hsengl totscore;

proc iml;
  title3 'Proportion of remaining variation explained by Pre-calculus';
  print "totscore controlling for hsgpa hscalcalc hsengl";
  n = 289; p = 5; s = 1; t = 3.92 ; F = t**2;
  a = s*F/(n-p + s*F); print a;
  print "For prediction, I am happy with Model 8:
  hsgpa hscalcalc hsengl totscore";

proc reg;
  title3 'Model 9: Same as Model 8 but including Mother Tongue';
  model grade = hsgpa hscalcalc hsengl totscore mtongue;

proc reg;
  title3 'Try automatic (stepwise) selection';
  model grade = hsgpa hscalcalc hsengl precalc calc
             mtongue gender e1-e4 e6
             / selection = stepwise slentry = 0.05 slstay = 0.05 ;
             /* Default slentry = slstay = 0.15 */

```

Skipping the checks of dummy variables ...

Prediction of Performance in First-year Calculus 5
 Variable Selection for Predicting Grade
 Model 1: Predict University Calculus Grade from HS Information

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

Number of Observations Read 579
 Number of Observations Used 323
 Number of Observations with Missing Values 256

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	45188	15063	73.23	<.0001
Error	319	65616	205.69147		
Corrected Total	322	110803			

Root MSE 14.34195 R-Square 0.4078
 Dependent Mean 59.79257 Adj R-Sq 0.4022
 Coeff Var 23.98617

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	Intercept	1	-77.07643	10.92142	-7.06	<.0001
hsgpa	High School GPA	1	1.76845	0.22097	8.00	<.0001
hscal	HS Calculus	1	0.30594	0.09750	3.14	0.0019
hsengl	HS English	1	-0.38461	0.11815	-3.26	0.0013

Prediction of Performance in First-year Calculus 6
 Variable Selection for Predicting Grade
 Proportion of remaining variation for HS information

hsgpa controlling for hscal and hsengl

a

0.1671018

hscal controlling for hsgpa and hsengl

a

0.0299812

hsengl controlling for hsgpa and hscal

a

0.0322412

Prediction of Performance in First-year Calculus
 Variable Selection for Predicting Grade
 Model 2: Predict University Calculus Grade from Diagnostic Test

7

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

Number of Observations Read 579
 Number of Observations Used 346
 Number of Observations with Missing Values 233

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	23658	11829	39.88	<.0001
Error	343	101730	296.58765		
Corrected Total	345	125388			

Root MSE 17.22172 R-Square 0.1887
 Dependent Mean 59.54335 Adj R-Sq 0.1840
 Coeff Var 28.92299

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value
Intercept	Intercept	1	37.85833	2.84884	13.29
precalc	Number precalculus correct	1	3.11970	0.63387	4.92
calc	Number calculus correct	1	1.80048	0.41033	4.39

Parameter Estimates

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

Prediction of Performance in First-year Calculus
 Variable Selection for Predicting Grade
 Model 3: Do the diagnostic test and HS info both contribute?

8

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

Number of Observations Read 579
 Number of Observations Used 289
 Number of Observations with Missing Values 290

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	44178	8835.57302	47.36	<.0001
Error	283	52793	186.54641		
Corrected Total	288	96970			

Root MSE 13.65820 R-Square 0.4556
 Dependent Mean 60.57785 Adj R-Sq 0.4460
 Coeff Var 22.54652

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value
Intercept	Intercept	1	-71.41781	11.14193	-6.41
hsgpa	High School GPA	1	1.59051	0.22122	7.19
hscal	HS Calculus	1	0.25421	0.10166	2.50
hsengl	HS English	1	-0.34853	0.12057	-2.89
precalc	Number precalculus correct	1	1.53973	0.55954	2.75
calc	Number calculus correct	1	0.68279	0.37603	1.82

Parameter Estimates

Variable	Label	DF	Pr > t
Intercept	Intercept	1	<.0001
hsgpa	High School GPA	1	<.0001
hscal	HS Calculus	1	0.0130
hsengl	HS English	1	0.0041
precalc	Number precalculus correct	1	0.0063
calc	Number calculus correct	1	0.0705

Prediction of Performance in First-year Calculus 9
 Variable Selection for Predicting Grade
 Model 3: Do the diagnostic test and HS info both contribute?

The REG Procedure
 Model: MODEL1

Test Diagnostic_Test Results for Dependent Variable grade

Source	DF	Mean Square	F Value	Pr > F
Numerator	2	1544.04857	8.28	0.0003
Denominator	283	186.54641		

Prediction of Performance in First-year Calculus 10
 Variable Selection for Predicting Grade
 Model 3: Do the diagnostic test and HS info both contribute?

The REG Procedure
 Model: MODEL1

Test HS_Information Results for Dependent Variable grade

Source	DF	Mean Square	F Value	Pr > F
Numerator	3	8761.88810	46.97	<.0001
Denominator	283	186.54641		

Prediction of Performance in First-year Calculus 11
 Variable Selection for Predicting Grade
 Proportion of remaining variation explained by diagnostic test

Precalc and calc controlling for hsgpa hscalc hsengl

a

0.0552811

Prediction of Performance in First-year Calculus
 Variable Selection for Predicting Grade
 Model 4: See if Course makes a contribution

12

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

Number of Observations Read 579
 Number of Observations Used 289
 Number of Observations with Missing Values 290

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	44713	6387.53844	34.35	<.0001
Error	281	52258	185.97057		
Corrected Total	288	96970			

Root MSE 13.63710 R-Square 0.4611
 Dependent Mean 60.57785 Adj R-Sq 0.4477
 Coeff Var 22.51170

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value
Intercept	Intercept	1	-72.42495	11.28800	-6.42
hsgpa	High School GPA	1	1.57595	0.22105	7.13
hscalc	HS Calculus	1	0.28249	0.10294	2.74
hsengl	HS English	1	-0.35419	0.12145	-2.92
precalc	Number precalculus correct	1	1.60358	0.56058	2.86
calc	Number calculus correct	1	0.73326	0.37721	1.94
c1	Catch-up	1	5.24279	4.35514	1.20
c3	Elite	1	-3.17149	2.78087	-1.14

Parameter Estimates

Variable	Label	DF	Pr > t
Intercept	Intercept	1	<.0001
hsgpa	High School GPA	1	<.0001
hscalc	HS Calculus	1	0.0065
hsengl	HS English	1	0.0038
precalc	Number precalculus correct	1	0.0045
calc	Number calculus correct	1	0.0529
c1	Catch-up	1	0.2297
c3	Elite	1	0.2551

Prediction of Performance in First-year Calculus 13
 Variable Selection for Predicting Grade
 Model 4: See if Course makes a contribution

The REG Procedure
 Model: MODEL1

Test Course Results for Dependent Variable grade

Source	DF	Mean Square	F Value	Pr > F
Numerator	2	267.45197	1.44	0.2391
Denominator	281	185.97057		

Prediction of Performance in First-year Calculus 14
 Variable Selection for Predicting Grade
 Model 4: See if Course makes a contribution

The REG Procedure
 Model: MODEL1

Test Diagnostic_Test Results for Dependent Variable grade

Source	DF	Mean Square	F Value	Pr > F
Numerator	2	1684.57843	9.06	0.0002
Denominator	281	185.97057		

Prediction of Performance in First-year Calculus
 Variable Selection for Predicting Grade
 Model 4 again with proc glm

16

The GLM Procedure

Dependent Variable: grade Final mark (if any)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	44712.76907	6387.53844	34.35	<.0001
Error	281	52257.72920	185.97057		
Corrected Total	288	96970.49827			

R-Square	Coeff Var	Root MSE	grade Mean
0.461097	22.51170	13.63710	60.57785

Source	DF	Type I SS	Mean Square	F Value	Pr > F
hsgpa	1	34512.00462	34512.00462	185.58	<.0001
hscalc	1	4744.09530	4744.09530	25.51	<.0001
hsengl	1	1833.66807	1833.66807	9.86	0.0019
precalc	1	2473.04238	2473.04238	13.30	0.0003
calc	1	615.05475	615.05475	3.31	0.0700
course	2	534.90395	267.45197	1.44	0.2391

Source	DF	Type III SS	Mean Square	F Value	Pr > F
hsgpa	1	9452.626595	9452.626595	50.83	<.0001
hscalc	1	1400.568197	1400.568197	7.53	0.0065
hsengl	1	1581.653505	1581.653505	8.50	0.0038
precalc	1	1521.786932	1521.786932	8.18	0.0045
calc	1	702.730145	702.730145	3.78	0.0529
course	2	534.903948	267.451974	1.44	0.2391

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Replicate Test of Course	2	534.903948	267.451974	1.44	0.2391
Diagnostic Test F = 9.06	2	3369.156864	1684.578432	9.06	0.0002

Prediction of Performance in First-year Calculus
 Variable Selection for Predicting Grade
 Model 5: Include Language, Sex and Ethnic Background

17

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

Number of Observations Read 579
 Number of Observations Used 287
 Number of Observations with Missing Values 292

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	12	46396	3866.29446	21.01	<.0001
Error	274	50434	184.06417		
Corrected Total	286	96829			

Root MSE 13.56702 R-Square 0.4791
 Dependent Mean 60.59582 Adj R-Sq 0.4563
 Coeff Var 22.38937

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value
Intercept	Intercept	1	-60.53085	11.87140	-5.10
hsgpa	High School GPA	1	1.57906	0.22245	7.10
hscal	HS Calculus	1	0.22034	0.10273	2.14
hsengl	HS English	1	-0.36429	0.12827	-2.84
precalc	Number precalculus correct	1	1.74734	0.56723	3.08
calc	Number calculus correct	1	0.65177	0.38352	1.70
mtongue	English vs. Other	1	-4.76328	2.19585	-2.17
gender		1	1.89029	1.72632	1.09
e1	Asian vs East Ind.	1	-5.49703	2.77555	-1.98
e2	East Eur. vs East Ind.	1	-5.46550	3.21790	-1.70
e3	Other Eur. vs East Ind.	1	-4.09079	2.54033	-1.61
e4	Mid. East & Pak. vs East Ind.	1	-5.98768	3.23591	-1.85
e6	Other/DK vs East Ind.	1	-1.05791	5.63847	-0.19

Parameter Estimates

Variable	Label	DF	Pr > t
Intercept	Intercept	1	<.0001
hsgpa	High School GPA	1	<.0001
hscal	HS Calculus	1	0.0328
hsengl	HS English	1	0.0048
precalc	Number precalculus correct	1	0.0023
calc	Number calculus correct	1	0.0904
mtongue	English vs. Other	1	0.0309
gender		1	0.2745
e1	Asian vs East Ind.	1	0.0486
e2	East Eur. vs East Ind.	1	0.0906
e3	Other Eur. vs East Ind.	1	0.1085
e4	Mid. East & Pak. vs East Ind.	1	0.0653
e6	Other/DK vs East Ind.	1	0.8513

Prediction of Performance in First-year Calculus 18
 Variable Selection for Predicting Grade
 Model 5: Include Language, Sex and Ethnic Background

The REG Procedure
 Model: MODEL1

Test TroubleVars Results for Dependent Variable grade

Source	DF	Mean Square	F Value	Pr > F
Numerator	7	305.41011	1.66	0.1191
Denominator	274	184.06417		

Prediction of Performance in First-year Calculus 19
 Variable Selection for Predicting Grade
 Model 5: Include Language, Sex and Ethnic Background

The REG Procedure
 Model: MODEL1

Test Nationality Results for Dependent Variable grade

Source	DF	Mean Square	F Value	Pr > F
Numerator	5	206.04701	1.12	0.3502
Denominator	274	184.06417		

Prediction of Performance in First-year Calculus 20
 Variable Selection for Predicting Grade
 Model 6: Discarding Gender and Nationality

The REG Procedure
 Model: MODEL1

Dependent Variable: grade Final mark (if any)

Number of Observations Read	579
Number of Observations Used	287
Number of Observations with Missing Values	292

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	45177	7529.47663	40.82	<.0001
Error	280	51652	184.47234		
Corrected Total	286	96829			

Root MSE	13.58206	R-Square	0.4666
Dependent Mean	60.59582	Adj R-Sq	0.4551
Coeff Var	22.41419		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value
Intercept	Intercept	1	-67.81236	11.26953	-6.02
hsgpa	High School GPA	1	1.57276	0.22101	7.12
hscal	HS Calculus	1	0.22495	0.10213	2.20
hsengl	HS English	1	-0.29798	0.12258	-2.43
precalc	Number precalculus correct	1	1.60480	0.55970	2.87
calc	Number calculus correct	1	0.61933	0.37553	1.65
mtongue	English vs. Other	1	-4.74223	2.12444	-2.23

Parameter Estimates

Variable	Label	DF	Pr > t
Intercept	Intercept	1	<.0001
hsgpa	High School GPA	1	<.0001
hscal	HS Calculus	1	0.0284
hsengl	HS English	1	0.0157
precalc	Number precalculus correct	1	0.0045
calc	Number calculus correct	1	0.1002
mtongue	English vs. Other	1	0.0264

Prediction of Performance in First-year Calculus 21
 Variable Selection for Predicting Grade
 Model 6: Discarding Gender and Nationality

The REG Procedure
 Model: MODEL1

Test EnglishTongue Results for Dependent Variable grade

Source	DF	Mean Square	F Value	Pr > F
Numerator	2	1253.11163	6.79	0.0013
Denominator	280	184.47234		

Prediction of Performance in First-year Calculus 22
 Variable Selection for Predicting Grade
 Proportion of remaining variation explained by mother tongue

Mtongue controlling for hsgpa hscal hsengl precalc calc

a

0.0174504

Prediction of Performance in First-year Calculus 23
 Variable Selection for Predicting Grade
 Model 7: Drop mtongue and calc
 Compare R-Square = 0.4556, Adj R-Sq = 0.4460 From Model 3

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

Number of Observations Read 579
 Number of Observations Used 289
 Number of Observations with Missing Values 290

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	43563	10891	57.91	<.0001
Error	284	53408	188.05524		
Corrected Total	288	96970			

Root MSE 13.71332 R-Square 0.4492
 Dependent Mean 60.57785 Adj R-Sq 0.4415
 Coeff Var 22.63752

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value
Intercept	Intercept	1	-74.11274	11.08720	-6.68
hsgpa	High School GPA	1	1.60015	0.22205	7.21
hscal	HS Calculus	1	0.29640	0.09937	2.98
hsengl	HS English	1	-0.35439	0.12101	-2.93
precalc	Number precalculus correct	1	1.90277	0.52470	3.63

Parameter Estimates

Variable	Label	DF	Pr > t
Intercept	Intercept	1	<.0001
hsgpa	High School GPA	1	<.0001
hscal	HS Calculus	1	0.0031
hsengl	HS English	1	0.0037
precalc	Number precalculus correct	1	0.0003

Prediction of Performance in First-year Calculus 24
 Variable Selection for Predicting Grade
 Proportion of remaining variation explained by Pre-calculus

precalc controlling for hsgpa hscalcl hsengl

a

0.0443403

Prediction of Performance in First-year Calculus 25
 Variable Selection for Predicting Grade
 Model 8: Combine precalc and calc instead of dropping calc
 Compare R-Square = 0.4492 from Model 7

The REG Procedure

Model: MODEL1

Dependent Variable: grade Final mark (if any)

Number of Observations Read 579
 Number of Observations Used 289
 Number of Observations with Missing Values 290

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	43951	10988	58.86	<.0001
Error	284	53019	186.68704		
Corrected Total	288	96970			

Root MSE 13.66335 R-Square 0.4532
 Dependent Mean 60.57785 Adj R-Sq 0.4455
 Coeff Var 22.55502

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error
Intercept	Intercept	1	-70.49310	11.11446
hsgpa	High School GPA	1	1.60612	0.22085
hscalcl	HS Calculus	1	0.24685	0.10148
hsengl	HS English	1	-0.35057	0.12060
totscore	Total # right on diagnostic test	1	0.98964	0.25277

Parameter Estimates

Variable	Label	DF	t Value	Pr > t
Intercept	Intercept	1	-6.34	<.0001
hsgpa	High School GPA	1	7.27	<.0001
hscalcl	HS Calculus	1	2.43	0.0156
hsengl	HS English	1	-2.91	0.0039
totscore	Total # right on diagnostic test	1	3.92	0.0001

Prediction of Performance in First-year Calculus 26
 Variable Selection for Predicting Grade
 Proportion of remaining variation explained by Pre-calculus
 totscore controlling for hsgpa hscalc hsengl

a

0.0513297

For prediction, I am happy with Model 8: hsgpa hscalc hsengl totscore

Prediction of Performance in First-year Calculus 27
 Variable Selection for Predicting Grade
 Model 9: Same as Model 8 but including Mother Tongue

The REG Procedure

Model: MODEL1

Dependent Variable: grade Final mark (if any)

Number of Observations Read	579
Number of Observations Used	287
Number of Observations with Missing Values	292

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	44881	8976.29511	48.56	<.0001
Error	281	51948	184.86704		
Corrected Total	286	96829			

Root MSE	13.59658	R-Square	0.4635
Dependent Mean	60.59582	Adj R-Sq	0.4540
Coeff Var	22.43815		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error
Intercept	Intercept	1	-66.75516	11.25053
hsgpa	High School GPA	1	1.58918	0.22087
hscalc	HS Calculus	1	0.21759	0.10208
hsengl	HS English	1	-0.30024	0.12270
totscore	Total # right on diagnostic test	1	0.97213	0.25185
mtongue	English vs. Other	1	-4.69657	2.12640

Parameter Estimates

Variable	Label	DF	t Value	Pr > t
Intercept	Intercept	1	-5.93	<.0001
hsgpa	High School GPA	1	7.20	<.0001
hscalc	HS Calculus	1	2.13	0.0339
hsengl	HS English	1	-2.45	0.0150
totscore	Total # right on diagnostic test	1	3.86	0.0001
mtongue	English vs. Other	1	-2.21	0.0280

Again, for prediction, I am happy with Model 8: hsgpa hscalc hsengl totscore. But mtongue is interesting. (Not strong, but interesting.)

But try an automatic method to see if it catches something we missed.

Prediction of Performance in First-year Calculus
 Variable Selection for Predicting Grade
 Try automatic (stepwise) selection

28

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

Number of Observations Read 579
 Number of Observations Used 287
 Number of Observations with Missing Values 292

Stepwise Selection: Step 1

Variable hsgpa Entered: R-Square = 0.3579 and C(p) = 54.7774

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	34656	34656	158.87	<.0001
Error	285	62173	218.14988		
Corrected Total	286	96829			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-87.53291	11.78465	12035	55.17	<.0001
hsgpa	1.82997	0.14519	34656	158.87	<.0001

Bounds on condition number: 1, 1

Stepwise Selection: Step 2

Variable hscalcalc Entered: R-Square = 0.4053 and C(p) = 31.8742

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	39240	19620	96.76	<.0001
Error	284	57589	202.77793		
Corrected Total	286	96829			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-78.92940	11.50506	9543.77385	47.07	<.0001
hsgpa	1.28138	0.18140	10118	49.90	<.0001
hscalcalc	0.45656	0.09603	4583.78372	22.60	<.0001

Bounds on condition number: 1.6795, 6.7178

Stepwise Selection: Step 3

Variable precalc Entered: R-Square = 0.4344 and C(p) = 18.5640

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	42058	14019	72.44	<.0001
Error	283	54771	193.53665		
Corrected Total	286	96829			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-77.39176	11.24707	9163.77010	47.35	<.0001
hsgpa	1.20854	0.17825	8896.80932	45.97	<.0001
hscalc	0.38611	0.09561	3156.10670	16.31	<.0001
precalc	2.03629	0.53364	2818.06020	14.56	0.0002

Bounds on condition number: 1.7445, 13.72

Stepwise Selection: Step 4

Variable hsengl Entered: R-Square = 0.4514 and C(p) = 11.6202

Prediction of Performance in First-year Calculus 29
 Variable Selection for Predicting Grade
 Try automatic (stepwise) selection

The REG Procedure

Model: MODEL1

Dependent Variable: grade Final mark (if any)

Stepwise Selection: Step 4

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	43704	10926	58.00	<.0001
Error	282	53125	188.38526		
Corrected Total	286	96829			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-74.38595	11.14286	8395.27582	44.56	<.0001
hsgpa	1.61314	0.22284	9871.54525	52.40	<.0001
hscalc	0.28738	0.10007	1553.64075	8.25	0.0044
hsengl	-0.35871	0.12135	1646.22898	8.74	0.0034
precalc	1.95702	0.52717	2596.17989	13.78	0.0002

Bounds on condition number: 2.7281, 29.826

Stepwise Selection: Step 5

Variable mtongue Entered: R-Square = 0.4614 and C(p) = 8.3468

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	44675	8935.02380	48.14	<.0001
Error	281	52154	185.60141		
Corrected Total	286	96829			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-70.13523	11.21533	7258.21332	39.11	<.0001
hsgpa	1.58158	0.22162	9452.36264	50.93	<.0001
hscal	0.26190	0.09995	1274.37374	6.87	0.0093
hsengl	-0.30256	0.12292	1124.43880	6.06	0.0144
precalc	1.93902	0.52332	2548.06067	13.73	0.0003
mtongue	-4.86990	2.12951	970.64664	5.23	0.0229

Bounds on condition number: 2.7387, 43.231

All variables left in the model are significant at the 0.0500 level.

No other variable met the 0.0500 significance level for entry into the model.

Summary of Stepwise Selection

Step	Variable Entered	Variable Removed	Label	Number Vars In	Partial R-Square
1	hsgpa		High School GPA	1	0.3579
2	hscal		HS Calculus	2	0.0473
3	precalc		Number precalculus correct	3	0.0291
4	hsengl		HS English	4	0.0170
5	mtongue		English vs. Other	5	0.0100

Summary of Stepwise Selection

Step	Model R-Square	C(p)	F Value	Pr > F
1	0.3579	54.7774	158.87	<.0001
2	0.4053	31.8742	22.60	<.0001
3	0.4344	18.5640	14.56	0.0002
4	0.4514	11.6202	8.74	0.0034
5	0.4614	8.3468	5.23	0.0229

This handout was prepared by Jerry Brunner, Department of Statistics, University of Toronto. It is licensed under a Creative Commons Attribution - ShareAlike 3.0 Unported License. Use any part of it as you like and share the result freely. The OpenOffice.org document is available from the course website:

<http://www.utstat.toronto.edu/~brunner/oldclass/appliedf13>