

# Sta442/1008f05 Overheads 6: Regression II

## Stepwise variable selection

---

Recall Jerry's favorite model for comparison. Try to beat this.

Variable	Label	DF	t Value	Pr >  t	Type I SS
Intercept	Intercept	1	-5.93	<.0001	1053822
gpa	High School GPA	1	7.20	<.0001	34656
english	Mark in HS English	1	-2.45	0.0150	4101.93295
hscalc	Mark in HS Calculus	1	2.13	0.0339	2349.96005
totscore	Total # right on diagnostic test	1	3.86	0.0001	2871.34040
tongue	Mother Tongue	1	-2.21	0.0280	901.84272

R-Square = 0.4635

---

```
/* mathstep1.sas */
%include 'mathexread3.sas';
title2 'Automatic variable selection: Goal is to predict grade';

options pagesize=1000; /* Print fewer page headings */

proc stepwise;
  title3 'Forward selection';
  model grade = gpa english hscalc credfm credag credcalc
    precalc1 precalc2 calcone calctwo precalc calc totscore
    tongue e1-e5 gender
    c1 c3
  / forward slentry=0.05;
  /* Default signif level for entry is 0.5 */

proc stepwise;
  title3 'Stepwise (mixed) selection';
  model grade = gpa english hscalc credfm credag credcalc
    precalc1 precalc2 calcone calctwo precalc calc totscore
    tongue e1-e5 gender
    c1 c3
  / stepwise slentry=0.05 slstay=0.05;
  /* Default slentry = slstay = 0.15 */

proc reg;
  title3 'Politically safe prediction model';
  model grade = gpa english hscalc totscore / ss1;
```

Math Diagnostic Study: Exploratory data 2  
 Automatic variable selection: Goal is to predict grade  
 Forward selection 21:43 Monday, October 10, 2005

The STEPWISE Procedure  
 Model: MODEL1  
 Dependent Variable: grade Mark in university calculus

Forward Selection: Step 1

Variable gpa Entered: R-Square = 0.3579 and C(p) = 50.8716

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
gpa	1.82997	0.14519	34656	158.87	<.0001

Bounds on condition number: 1, 1

Forward Selection: Step 2

Variable totscore Entered: R-Square = 0.4144 and C(p) = 23.4790

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	40130	20065	100.50	<.0001
Error	284	56699	199.64543		
Corrected Total	286	96829			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-77.92562	11.42210	9292.39457	46.54	<.0001
gpa	1.56942	0.14754	22591	113.15	<.0001
totscore	1.30211	0.24868	5473.41474	27.42	<.0001

Bounds on condition number: 1.1284, 4.5134

Forward Selection: Step 3

Variable english Entered: R-Square = 0.4435 and C(p) = 10.3756

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	42942	14314	75.17	<.0001
Error	283	53887	190.41272		
Corrected Total	286	96829			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-72.85566	11.23260	8010.54144	42.07	<.0001
gpa	1.94618	0.17427	23747	124.71	<.0001
english	-0.44468	0.11570	2812.50227	14.77	0.0001
totscore	1.15290	0.24595	4183.98406	21.97	<.0001

Bounds on condition number: 1.6507, 12.829

Forward Selection: Step 4

Variable tongue Entered: R-Square = 0.4548 and C(p) = 6.4730

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	44041	11010	58.82	<.0001
Error	282	52788	187.19014		
Corrected Total	286	96829			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-68.22951	11.29959	6824.99653	36.46	<.0001
gpa	1.87948	0.17497	21599	115.38	<.0001
english	-0.37609	0.11816	1896.30271	10.13	0.0016
totscore	1.11430	0.24438	3891.84700	20.79	<.0001
tongue	-5.15807	2.12860	1099.17931	5.87	0.0160

Bounds on condition number: 1.6926, 21.954

Forward Selection: Step 5

Variable hscalC Entered: R-Square = 0.4635 and C(p) = 3.9622

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			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-66.75516	11.25053	6508.53209	35.21	<.0001
gpa	1.58918	0.22087	9570.77619	51.77	<.0001
english	-0.30024	0.12270	1106.96199	5.99	0.0150
hscalC	0.21759	0.10208	839.97980	4.54	0.0339
totscore	0.97213	0.25185	2754.41724	14.90	0.0001
tongue	-4.69657	2.12640	901.84272	4.88	0.0280

Bounds on condition number: 2.7308, 44.252

Forward Selection: Step 6

Variable e4 Entered: R-Square = 0.4710 and C(p) = 2.0600

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	45608	7601.35727	41.55	<.0001
Error	280	51221	182.93204		
Corrected Total	286	96829			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-64.60808	11.24323	6040.62097	33.02	<.0001
gpa	1.58717	0.21971	9546.34142	52.19	<.0001
english	-0.32445	0.12266	1279.97837	7.00	0.0086
hscalC	0.20766	0.10166	763.23017	4.17	0.0420
totscore	0.95113	0.25075	2632.00623	14.39	0.0002
tongue	-4.48110	2.11801	818.85180	4.48	0.0353
e4	4.57904	2.29748	726.66807	3.97	0.0472

Bounds on condition number: 2.7309, 59.472

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

Summary of Forward Selection

Step	Variable Entered	Label	Number Vars In	Partial R-Square	Model R-Square
1	gpa	High School GPA	1	0.3579	0.3579
2	totscore	Total # right on diagnostic test	2	0.0565	0.4144
3	english	Mark in HS English	3	0.0290	0.4435
4	tongue	Mother Tongue	4	0.0114	0.4548
5	hscal	Mark in HS Calculus	5	0.0087	0.4635
6	e4		6	0.0075	0.4710

Summary of Forward Selection

Step	C(p)	F Value	Pr > F
1	50.8716	158.87	<.0001
2	23.4790	27.42	<.0001
3	10.3756	14.77	0.0001
4	6.4730	5.87	0.0160
5	3.9622	4.54	0.0339
6	2.0600	3.97	0.0472

---

Math Diagnostic Study: Exploratory data 3  
 Automatic variable selection: Goal is to predict grade  
 Stepwise (mixed) selection  
 21:43 Monday, October 10, 2005

The STEPWISE Procedure  
 Model: MODEL1  
 Dependent Variable: grade Mark in university calculus

Stepwise Selection: Step 1

Variable gpa Entered: R-Square = 0.3579 and C(p) = 50.8716

There rest of the output for stepwise is the same as forward, because no variables met the criterion for removal.

Politically Safe Model	R-Square = 0.4532
Jerry's Model	R-Square = 0.4635
Stepwise Model	R-Square = 0.4710