

Variable Selection

```
/* smsa2.sas */
options linesize=79 pagesize=100;
title 'STA 302 Summer 2001: SMSA Data';
title2 'Automatic variable selection';

data census;
  infile 'smsa.dat';
  input id landarea totpop urban oldfolks doctors hospbeds hsgrads
        labforce income crimes region;
  if region=2 then r1=1; else r1=0;
  if region=3 then r2=1; else r2=0;
  if region=4 then r3=1; else r3=0;
  crimrate = crimes/totpop;
  density = totpop/landarea;
  labrate = labforce/totpop;
  aveinc = income/totpop;

proc corr; /* Correlations of all variables with each other */
  var crimrate landarea -- income density labrate aveinc r1 r2 r3;

proc reg; /* Try all variables at once */
  model crimrate = landarea -- income density labrate aveinc;

proc reg;
  model crimrate = landarea -- income density labrate aveinc /
    selection = rsquare best=3;

proc reg;
  model crimrate = landarea urban doctors hospbeds hsgrads / ss1;
  model crimrate = landarea urban doctors hospbeds hsgrads
    density aveinc totpop/ ss1;
  xtra: test density=aveinc=totpop = 0;
  model crimrate = landarea urban doctors hospbeds hsgrads r1 r2 r3/ ss1;
  region: test r1=r2=r3=0;

proc reg;
  model crimrate = r1 r2 r3 landarea -- income density labrate aveinc /
    selection = rsquare include=3 best=3;
  /* Always include region */

proc reg;
  model crimrate = r1 r2 r3 landarea hsgrads density aveinc;
```

```
proc reg;
  model crimrate = landarea -- income density labrate aveinc /
    selection = rsquare best=3;
```

The REG Procedure
 Model: MODEL1
 Dependent Variable: crimrate

R-Square Selection Method

Number in Model	R-Square	Variables in Model
1	0.1413	hsgrads
1	0.1182	landarea
1	0.0495	urban

2	0.2248	landarea hsgrads
2	0.1753	urban hsgrads
2	0.1735	doctors hsgrads

3	0.2615	landarea urban hsgrads
3	0.2484	landarea doctors hsgrads
3	0.2435	landarea hsgrads density

4	0.2914	landarea doctors hospbeds hsgrads
4	0.2838	landarea urban doctors hsgrads
4	0.2791	landarea urban hsgrads income

5	0.3205	landarea urban doctors hospbeds hsgrads
5	0.2957	landarea totpop urban hospbeds hsgrads
5	0.2954	landarea doctors hospbeds hsgrads density

6	0.3244	landarea urban doctors hospbeds hsgrads density
6	0.3242	landarea urban doctors hospbeds hsgrads aveinc
6	0.3223	landarea totpop urban doctors hospbeds hsgrads

7	0.3284	landarea urban doctors hospbeds hsgrads labrate aveinc
7	0.3266	landarea urban doctors hospbeds hsgrads density aveinc
7	0.3263	landarea totpop urban doctors hospbeds hsgrads density

8	0.3302	landarea totpop urban doctors hospbeds hsgrads labforce aveinc
8	0.3300	landarea urban doctors hospbeds hsgrads density labrate aveinc
8	0.3295	landarea totpop urban doctors hospbeds hsgrads income aveinc

9	0.3340	landarea	totpop	urban	doctors	hospbeds	hsgrads	income
								labrate
								aveinc
9	0.3324	landarea	totpop	urban	doctors	hospbeds	hsgrads	labforce
								income
								aveinc
9	0.3322	landarea	totpop	urban	doctors	hospbeds	hsgrads	income
								density
								aveinc

10	0.3359	landarea	totpop	urban	doctors	hospbeds	hsgrads	income
								density
								labrate
								aveinc
10	0.3342	landarea	totpop	urban	doctors	hospbeds	hsgrads	labforce
								income
								density
								aveinc
10	0.3342	landarea	totpop	urban	oldfolks	doctors	hospbeds	hsgrads
								income
								labrate
								aveinc

11	0.3361	landarea	totpop	urban	oldfolks	doctors	hospbeds	hsgrads
								income
								density
								labrate
								aveinc
11	0.3359	landarea	totpop	urban	doctors	hospbeds	hsgrads	labforce
								income
								density
								labrate
								aveinc
11	0.3345	landarea	totpop	urban	oldfolks	doctors	hospbeds	hsgrads
								labforce
								income
								density
								aveinc

12	0.3362	landarea	totpop	urban	oldfolks	doctors	hospbeds	hsgrads
								labforce
								income
								density
								labrate
								aveinc

Picking up with test of "extra" variables density aveinc & totpop, ...

xtra: test density=aveinc=totpop = 0;

Test XTRA Results for Dependent Variable crimrate

Source	DF	Mean Square	F Value	Pr > F
Numerator	3	71.17226	0.48	0.6973
Denominator	132	148.51288		

Now test region

```
model crimrate = landarea urban doctors hospbeds hsgrads r1 r2 r3/ ssl;
region: test r1=r2=r3=0;
```

The REG Procedure
 Model: MODEL3
 Dependent Variable: crimrate

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	13922	1740.25551	15.07	<.0001
Error	132	15244	115.48407		
Corrected Total	140	29166			

Root MSE	10.74635	R-Square	0.4773
Dependent Mean	55.86704	Adj R-Sq	0.4457
Coeff Var	19.23559		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Type I SS
Intercept	1	19.82563	8.15669	2.43	0.0164	440079
landarea	1	0.00066215	0.00036716	1.80	0.0736	3447.74942
urban	1	0.02130	0.05603	0.38	0.7044	1477.28111
doctors	1	0.00195	0.00143	1.36	0.1765	868.41784
hospbeds	1	-0.00032785	0.00050830	-0.64	0.5200	1974.32292
hsgrads	1	0.36589	0.14728	2.48	0.0142	1580.95289
r1	1	8.91097	2.98332	2.99	0.0034	311.38036
r2	1	15.81309	2.85891	5.53	<.0001	1233.68682
r3	1	20.65661	4.03389	5.12	<.0001	3028.25269

Test REGION Results for Dependent Variable crimrate

Source	DF	Mean Square	F Value	Pr > F
Numerator	3	1524.43996	13.20	<.0001
Denominator	132	115.48407		

Clearly want region in the model. Along with what?

```

proc reg;
  model crimrate = r1 r2 r3 landarea -- income density labrate aveinc /
    selection = rsquare include=3 best=3;
    /* Always include region */

```

The REG Procedure
 Model: MODEL1
 Dependent Variable: crimrate

R-Square Selection Method

NOTE: The variables in the 3 variable model are included in all models.

Number in Model	R-Square	Variables in Model
3	0.3825	r1 r2 r3
4	0.4576	aveinc
4	0.4404	doctors
4	0.4348	income
5	0.4772	landarea aveinc
5	0.4743	density aveinc
5	0.4726	doctors aveinc
6	0.4981	landarea density aveinc
6	0.4919	landarea hsgrads density
6	0.4890	hsgrads density aveinc
7	0.5159	landarea hsgrads density aveinc
7	0.5040	landarea doctors hsgrads density
7	0.5026	landarea urban density aveinc
8	0.5187	landarea doctors hsgrads density aveinc
8	0.5180	landarea urban hsgrads density aveinc
8	0.5178	landarea hospbeds hsgrads density aveinc
9	0.5297	landarea totpop hsgrads labforce density aveinc
9	0.5255	landarea doctors hsgrads labforce density aveinc
9	0.5247	landarea doctors hsgrads income density aveinc
10	0.5344	landarea totpop hsgrads labforce density labrate aveinc
10	0.5316	landarea totpop doctors hsgrads labforce density aveinc
10	0.5309	landarea totpop urban hsgrads labforce density aveinc
11	0.5361	landarea totpop hospbeds hsgrads labforce density labrate aveinc
11	0.5349	landarea totpop doctors hsgrads labforce density labrate aveinc
11	0.5347	landarea totpop urban hsgrads labforce density labrate aveinc

12	0.5376	landarea	totpop	doctors	hospbeds	hsgrads	labforce	density	labrate	aveinc			
12	0.5365	landarea	totpop	urban	hospbeds	hsgrads	labforce	density	labrate	aveinc			
12	0.5362	landarea	totpop	oldfolks	hospbeds	hsgrads	labforce	density	labrate	aveinc			

13	0.5382	landarea	totpop	doctors	hospbeds	hsgrads	labforce	income	density	labrate	aveinc		
13	0.5379	landarea	totpop	urban	doctors	hospbeds	hsgrads	labforce	density	labrate	aveinc		
13	0.5376	landarea	totpop	oldfolks	doctors	hospbeds	hsgrads	labforce	density	labrate	aveinc		

14	0.5385	landarea	totpop	urban	doctors	hospbeds	hsgrads	labforce	income	density	labrate	aveinc	
14	0.5383	landarea	totpop	oldfolks	doctors	hospbeds	hsgrads	labforce	income	density	labrate	aveinc	
14	0.5381	landarea	totpop	urban	oldfolks	doctors	hospbeds	hsgrads	labforce	density	labrate	aveinc	

15	0.5386	landarea	totpop	urban	oldfolks	doctors	hospbeds	hsgrads	labforce	income	density	labrate	aveinc

All the 7-variable models are promising. Look at the first one as an example.

```
proc reg;
  model crimrate = r1 r2 r3 landarea hsgrads density aveinc;
```

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	-5.59355	9.99325	-0.56	0.5766
r1	1	10.73255	2.78545	3.85	0.0002
r2	1	20.33952	2.68800	7.57	<.0001
r3	1	23.81440	3.49635	6.81	<.0001
landarea	1	0.00094151	0.00034615	2.72	0.0074
hsgrads	1	0.34415	0.15540	2.21	0.0285
density	1	2.46848	0.84455	2.92	0.0041
aveinc	1	3.77557	1.46813	2.57	0.0112