

Model Selection

```

DATA; INFILE squid;
INPUT x1 x2 x3 x4 x5 y;

PROC RSQUARE;
MODEL y=x1 x2 x3 x4 x5 / ADJRSQ CP SSE MSE;

PROC REG;
MODEL y=x1 x2 x3 x4 x5 / SELECTION=STEPWISE;

PROC REG;
model y=x1 x2 x3 x4 x5 / SELECTION=BACKWARD
      SLENTRY=0.05 SLSTAY=0.05;
RUN;

```

/* R-square etc. for all models */

/* Stepwise election of the best */

/* set of predictors */

/* Backward selection with */

/* specified significance levels */

The RSQUARE Procedure
R-Square Selection Method

Number in Model	R-Square	Adjusted R-Square	C(p)	MSE	SSE	Variables in Model
1	0.9455	0.9428	5.7794	0.58836	11.76713	x5
1	0.9223	0.9184	15.9084	0.83897	16.77941	x1
1	0.9140	0.9097	19.5335	0.92866	18.57326	x3
1	0.8857	0.8800	31.8621	1.23370	24.67401	x4
1	0.8812	0.8753	33.8244	1.28225	25.64506	x2

2	0.9584	0.9540	2.1456	0.47259	8.97925	x4 x5
2	0.9575	0.9530	2.5565	0.48329	9.18259	x1 x5
2	0.9486	0.9432	6.4172	0.58384	11.09305	x2 x5
2	0.9477	0.9422	6.8317	0.59464	11.29813	x3 x5
2	0.9386	0.9322	10.7846	0.69759	13.25422	x1 x3
2	0.9288	0.9213	15.0819	0.80951	15.38074	x3 x4
2	0.9229	0.9148	17.6524	0.87646	16.65275	x1 x2
2	0.9226	0.9145	17.7619	0.87931	16.70693	x1 x4
2	0.9167	0.9079	20.3553	0.94686	17.99026	x2 x3
2	0.8960	0.8850	29.3891	1.18213	22.46056	x2 x4
3	0.9616	0.9553	2.7354	0.46008	8.28144	x2 x4 x5
3	0.9592	0.9524	3.8138	0.48973	8.81505	x1 x4 x5
3	0.9585	0.9516	4.0951	0.49746	8.95427	x3 x4 x5
3	0.9584	0.9515	4.1525	0.49904	8.98266	x1 x2 x5
3	0.9575	0.9504	4.5563	0.51014	9.18251	x1 x3 x5
3	0.9488	0.9403	8.3265	0.61379	11.04816	x2 x3 x5
3	0.9420	0.9323	11.3275	0.69629	12.53321	x1 x2 x3
3	0.9386	0.9284	12.7821	0.73628	13.25299	x1 x3 x4
3	0.9314	0.9200	15.9325	0.82289	14.81195	x2 x3 x4
3	0.9229	0.9101	19.6254	0.92441	16.63935	x1 x2 x4

4	0.9630	0.9543	4.1582	0.47034	7.99580	x1 x2 x4 x5
4	0.9619	0.9530	4.6037	0.48331	8.21625	x2 x3 x4 x5
4	0.9593	0.9497	5.7556	0.51684	8.78628	x1 x3 x4 x5
4	0.9588	0.9491	5.9859	0.52354	8.90021	x1 x2 x3 x5
4	0.9432	0.9298	12.7951	0.72175	12.26972	x1 x2 x3 x4

5	0.9633	0.9519	6.0000	0.49485	7.91752	x1 x2 x3 x4 x5

The REG Procedure

Stepwise Selection: Step 1

Variable x5 Entered: R-Square = 0.9455 and C(p) = 5.7794

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	204.15762	204.15762	347.00	<.0001
Error	20	11.76713	0.58836		
Corrected Total	21	215.92475			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-5.48434	0.54474	59.63564	101.36	<.0001
x5	20.85657	1.11964	204.15762	347.00	<.0001

Stepwise Selection: Step 2
Variable x4 Entered: R-Square = 0.9584 and C(p) = 2.1456

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	206.94550	103.47275	218.95	<.0001
Error	19	8.97925	0.47259		
Corrected Total	21	215.92475			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-6.33511	0.60088	52.53154	111.16	<.0001
x4	4.15425	1.71041	2.78787	5.90	0.0252
x5	15.01600	2.60568	15.69476	33.21	<.0001

All variables left in the model are significant at the 0.1500 level.
No other variable met the 0.1500 significance level for entry into the model.

Summary of Stepwise Selection

Step	Variable Entered	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	x5		1	0.9455	0.9455	5.7794	347.00	<.0001
2	x4		2	0.0129	0.9584	2.1456	5.90	0.0252

The REG Procedure Backward Elimination: Step 0

All Variables Entered: R-Square = 0.9633 and C(p) = 6.0000

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	208.00723	41.60145	84.07	<.0001
Error	16	7.91752	0.49485		
Corrected Total	21	215.92475			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-6.51221	0.93356	24.07918	48.66	<.0001
x1	1.99941	2.57334	0.29873	0.60	0.4485
x2	-3.67510	2.77366	0.86876	1.76	0.2038
x3	2.52449	6.34750	0.07827	0.16	0.6961
x4	5.15808	3.66028	0.98269	1.99	0.1779
x5	14.40116	4.85599	4.35219	8.80	0.0091

Backward Elimination: Step 1
Variable x3 Removed: R-Square = 0.9630 and C(p) = 4.1582

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	207.92895	51.98224	110.52	<.0001
Error	17	7.99580	0.47034		

Corrected Total 21 215.92475

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-6.34844	0.81683	28.41077	60.40	<.0001
x1	1.95314	2.50625	0.28565	0.61	0.4465
x2	-3.16590	2.39880	0.81926	1.74	0.2044
x4	5.16889	3.56841	0.98687	2.10	0.1657
x5	15.63277	3.64667	8.64356	18.38	0.0005

Backward Elimination: Step 2

Variable x1 Removed: R-Square = 0.9616 and C(p) = 2.7354

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	207.64331	69.21444	150.44	<.0001
Error	18	8.28144	0.46008		
Corrected Total	21	215.92475			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-5.98038	0.65914	37.87384	82.32	<.0001
x2	-2.88979	2.34647	0.69781	1.52	0.2340
x4	6.87926	2.78280	2.81161	6.11	0.0236
x5	17.10900	3.08189	14.17911	30.82	<.0001

Backward Elimination: Step 3

Variable x2 Removed: R-Square = 0.9584 and C(p) = 2.1456

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	206.94550	103.47275	218.95	<.0001
Error	19	8.97925	0.47259		
Corrected Total	21	215.92475			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-6.33511	0.60088	52.53154	111.16	<.0001
x4	4.15425	1.71041	2.78787	5.90	0.0252
x5	15.01600	2.60568	15.69476	33.21	<.0001

All variables left in the model are significant at the 0.1000 level.

Summary of Backward Elimination

Step	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	x3	4	0.0004	0.9630	4.1582	0.16	0.6961
2	x1	3	0.0013	0.9616	2.7354	0.61	0.4465
3	x2	2	0.0032	0.9584	2.1456	1.52	0.2340