Variable Selection Methods

- PROBLEM: Find a set of predictor variables which gives a good fit, predicts the dependent value well and is as small as possible.
- So far have used F and t tests to compare 2 models at a time.
- Have followed a sequence of tests to try to find a good set of variables but our method has been informal
- Other statisticians using the same method might select a different final model.
- Now investigate 4 mechanical (more or less) variable selection methods: Forward, Backward, Stepwise and All Subsets.

Forward Selection

- Start with a model with no predictors.
- Add variable with largest F-statistic (provided P less than some cut-off).
- Refit with this variable. Recompute all F statistics for adding one of the remaining variables and add variable with largest F statistic.
- Continue until no variable is significant at cut-off level.

Backward Selection

- Start with model with all predictors.
- Delete variable with smallest F-statistic (provided P more than some cut-off).
- Refit with this variable deleted. Recompute all F statistics for deleting one of the remaining variables and delete variable with smallest F statistic.
- Continue until every remaining variable is significant at cut-off level.

Stepwise Selection

- Start with model with no predictors.
- Add variable with largest F-statistic (provided P less than some cut-off).
- Refit with this variable added. Recompute all F statistics for adding one of the remaining variables and add variable with largest F statistic.
- At each step after adding a variable try to eliminate any variable not significant at some level (that is, do BACKWARD elimination till that stops).
- After doing the backwards steps take another FORWARD step.
- Continue until every remaining variable is significant at cut-off level and every excluded variable is insignificant OR until variable to be added is same as last deleted variable.

All Subsets

- For each subset of the set of predictors fit the model and compute some summary statistic of the quality of the fit.
- Pick model which makes this summary as large (or sometimes as small) as possible.
- With k predictors fit 2^k models; impractical for k too large. Special **Best subsets** algorithms work without looking at all 2^k models.
- Possible summary statistics:
  - R^2: but NOTE — adding a variable increases R^2 so this is most useful for comparing models of the same size.
  - Adjusted R^2: This method adjusts R^2 to try to compensate for the fact that more variables produces larger R^2 even when the extra variables are irrelevant.
  - C_p: Like Adjusted R^2 but based on a trade off of bias and variance.
  - PRESS: The sum of squares of the PRESS residuals (See material on Diagnostics.)

Example: Forward Selection

```
data scenic;
 infile 'scenic.dat' firstobs=2;
 input Stay  Age Risk Culture Chest Beds
    School Region Census Nurses Facil;
 Nratio = Nurses / Census  ;
proc reg  data=scenic;
  model Risk = Culture Stay Nurses Nratio
      Chest Beds Census Facil /
      selection=forward;
run ;
```

(Complete output)

Edited SAS Output

```
   Forward Selection Proc for Dependent Variable RISK
Step 1 Var CULTURE Entered R-sq=0.3127 C(p)=47.48
            DF Sum Sq   Mean Sq   F   Prob>F
Regression   1   62.9631   62.9631  50.49 0.0001
Error      111  138.4167    1.2470
Total      112  201.37982301
           Par  Std   Type II
Variable    Est  Error    Sum Sq     F    Prob>F
INTERCEP  3.1979 0.1938  339.6491 272.37 0.0001
CULTURE   0.0733 0.0103   62.9631  50.49 0.0001
-----------------------------------------------
```

Edited SAS Output

```
Step 2 Var STAY Entered R-sq=0.450 C(p)=18.12
            DF Sum Sq Mean Sq   F   Prob>F
Regression   2  90.7020   45.3510 45.07 0.0001
Error      110 110.6778    1.0061
Total      112 201.37982301
           Par    Std Type II
Variable    Est      Error  Sum Sq     F Prob>F
INTERCEP  0.80549 0.48776  2.7440  2.73 0.1015
CULTURE   0.05645 0.00980 33.3969 33.19 0.0001
STAY      0.27547 0.05246 27.7388 27.57 0.0001
-----------------------------------------------
```

Edited SAS Output

```
Step 3 Var FACIL Entered R-sq=0.493 C(p)=10.33
            DF Sum of Sq Mean Sq    F   Prob>F
Regression   3   99.3608 33.1203 35.39 0.0001
Error      109  102.0190  0.9360
Total      112  201.3798
            Par   Std      Type II
Variable     Est  Error    Sum Sq     F  Prob>F
INTERCEP  0.4913 0.4816  0.9740  1.04 0.3099
CULTURE   0.0542 0.0095 30.5982 32.69 0.0001
STAY      0.2239 0.0534 16.4766 17.60 0.0001
FACIL     0.0196 0.0065  8.6588  9.25 0.0029
-----------------------------------------
```

Edited SAS Output

```
Step 4 Var NRATIO Entered R-sq=0.525 C(p)= 5.03
              DF Sum of Sq Mean Sq    F  Prob>F
Regression     4 105.8210 26.4552 29.90 0.0001
Error        108  95.5589  0.8848
Total        112 201.3798
              Par   Std       Type II
Variable      Est  Error    Sum Sq    F   Prob>F
INTERCEP   -0.4951 0.5938  0.6151  0.70 0.4063
CULTURE     0.0482 0.0095 22.8451 25.82 0.0001
STAY        0.2676 0.0543 21.4500 24.24 0.0001
NRATIO      0.7926 0.2933  6.4601  7.30 0.0080
FACIL       0.0175 0.0063  6.7535  7.63 0.0067
-----------------------------------------------
```

Edited SAS Output

```
Step 5 Var CHEST Entered R-sq=0.5379 C(p)= 4.195
                DF    Sum Sq   Mean Sq   F   Prob>F
Regression       5  108.3272   21.6654  24.91 0.0001
Error          107   93.0527    0.86965099
Total          112  201.37982301
              Par    Std      Type II
Variable      Est   Error    Sum Sq      F    Prob>F
INTERCEP  -0.7680 0.6102    1.3776    1.58 0.2109
CULTURE    0.0432 0.0098   16.7198   19.23 0.0001
STAY       0.2339 0.0574   14.4381   16.60 0.0001
NRATIO     0.6724 0.2993    4.3888    5.05 0.0267
CHEST      0.0092 0.0054    2.5062    2.88 0.0925
FACIL      0.0184 0.0063    7.4571    8.57 0.0042
-----------------------------------------------
```

Edited SAS Output

```
Step 6 Var CENSUS Entered R-sq=0.54147 C(p)=5.39
            DF Sum Sq  Mean Sq      F Prob>F
Regression   6   109.0408  18.1735 20.86 0.0001
Error      106    92.3390   0.87112288
Total      112   201.37982301
            Par      Std      Type II
Variable     Est        Error Sum Sq   F Prob>F
INTERCEP  -0.6098  0.6353   0.8028  0.92 0.3393
CULTURE    0.0433  0.0099  16.7860 19.27 0.0001
STAY       0.2181  0.0601  11.4796 13.18 0.0004
NRATIO     0.7425  0.3094   5.0165  5.76 0.0182
CHEST      0.0097  0.0054   2.7548  3.16 0.0782
CENSUS     0.0009  0.0010   0.7136  0.82 0.3675
FACIL      0.0117  0.0097   1.2598  1.45 0.2318
-----------------------------------------------
```

Edited SAS Output

No other variable met the 0.5000 significance level for entry into the model. Summary of Forward Seln Procedure for Dependent Variable RISK

Step	Variable Entered	Num In	Partl R**2	Model R**2	C(p)	F	Prob>F
1	CULTURE	1	0.3127	0.3127	47.4779	50.49	0.0001
2	STAY	2	0.1377	0.4504	18.1196	27.57	0.0001
3	FACIL	3	0.0430	0.4934	10.3309	9.25	0.0029
4	NRATIO	4	0.0321	0.5255	5.0278	7.30	0.0080
5	CHEST	5	0.0124	0.5379	4.1946	2.88	0.0925
6	CENSUS	6	0.0035	0.5415	5.3879	0.82	0.3675

Example:Backward Selection

```
data scenic;
 infile 'scenic.dat' firstobs=2;
 input Stay  Age Risk Culture Chest Beds
    School Region Census Nurses Facil;
 Nratio = Nurses / Census  ;
proc reg  data=scenic;
  model Risk = Culture Stay Nurses Nratio
      Chest Beds Census Facil /
      selection=backward;
run ;
```

(Complete output)

Edited SAS output

```
Backward Elimination Proc for Dependent Var RISK
Step 0 All Variables Entrd R-sq=0.543 C(p)=9.00
           DF Sum Sq  Mean Sq      F  Prob>F
Regression  8 109.3839 13.6730 15.46 0.0001
Error     104  91.9959  0.8846
Total     112 201.3798
          Par    Std    Type II
Variable   Est   Error  Sum Sq     F  Prob>F
INTERCEP -0.6154 0.6664  0.7544   0.85 0.3579
CULTURE   0.0441 0.0100 17.0508 19.28 0.0001
STAY      0.2054 0.0641  9.0980 10.29 0.0018
NURSES   -0.0009 0.0022  0.1453  0.16 0.6861
NRATIO    0.8501 0.3933  4.1320  4.67 0.0330
CHEST     0.0095 0.0055  2.6250  2.97 0.0879
BEDS     -0.0011 0.0027  0.1426  0.16 0.6888
CENSUS    0.0030 0.0036  0.6031  0.68 0.4109
FACIL     0.0131 0.0101  1.4914  1.69 0.1970
```

Edited SAS Output

```
Step 1 Var BEDS Removed R-sq=0.5425 C(p)=7.161249
              DF Sum Sq Mean Sq      F  Prob>F
Regression     7  109.2413  15.606  17.78 0.0001
Error        105   92.1386   0.87751018
Total        112  201.37982301
              Par    Std         Type II
Variable      Est      Error Sum Sq     F Prob>F
INTERCEP   -0.6699 0.6499    0.9325  1.06 0.3050
CULTURE     0.0440 0.0100   16.9645 19.33 0.0001
STAY        0.2122 0.0615   10.4433 11.90 0.0008
NURSES     -0.0010 0.0021    0.2005  0.23 0.6337
NRATIO      0.8564 0.3915    4.2003  4.79 0.0309
CHEST       0.0095 0.0055    2.6267  2.99 0.0865
CENSUS      0.0018 0.0021    0.6502  0.74 0.3913
FACIL       0.0123 0.0098    1.3651  1.56 0.2151
-----------------------------------------------
```

Edited SAS Output

```
Step 2 Var NURSES Removed R-sq=0.54 C(p)=5.39
           DF  Sum Sq  Mean Sq   F   Prob>F
Regression  6 109.0408 18.1735 20.86 0.0001
Error     106  92.3390  0.8711
Total     112 201.3798
             Par    Std  Type II
Variable     Est   Error Sum Sq     F  Prob>F
INTERCEP -0.6098 0.6353  0.8028  0.92 0.3393
CULTURE   0.0433 0.0099 16.7860 19.27 0.0001
STAY      0.2181 0.0601 11.4796 13.18 0.0004
NRATIO    0.7425 0.3094  5.0165  5.76 0.0182
CHEST     0.0097 0.0054  2.7548  3.16 0.0782
CENSUS    0.0009 0.0010  0.7136  0.82 0.3675
FACIL     0.0117 0.0097  1.2598  1.45 0.2318
-----------------------------------------
```

Edited SAS Output

```
Step 3 Var CENSUS Removed R-sq=0.538 C(p)=4.20
             DF  Sum Sq  Mean Sq   F  Prob>F
Regression    5 108.3272  21.67 24.91 0.0001
Error       107  93.0527   0.87
Total       112 201.37982301
            Par   Std    Type II
Variable    Est    Error  Sum Sq    F Prob>F
INTERCEP -0.7680 0.6102  1.3776  1.58 0.2109
CULTURE   0.0432 0.0098 16.7198 19.23 0.0001
STAY      0.2339 0.0574 14.4381 16.60 0.0001
NRATIO    0.6724 0.2993  4.3888  5.05 0.0267
CHEST     0.0092 0.0054  2.5062  2.88 0.0925
FACIL     0.0184 0.0063  7.4571  8.57 0.0042
-----------------------------------------------
```

```
All variables left in the model are
significant at the 0.1000 level.
Summary of Backward Elimination Proc
for Dependent Variable RISK
     Vare Num Partl  Model
Step Remd  In  R**2   R**2 C(p)     F  Prob>F
 1   BEDS   7  0.001 0.542 7.16 0.1612 0.6888
 2   NURSES 6  0.001 0.542 5.39 0.2284 0.6337
 3   CENSUS 5  0.004 0.538 4.19 0.8192 0.3675
```

Stepwise Selection

```
data scenic;
 infile 'scenic.dat' firstobs=2;
 input Stay  Age Risk Culture Chest Beds
     School Region Census Nurses Facil;
 Nratio = Nurses/Census;
proc reg  data=scenic;
  model Risk = Culture Stay Nurses Nratio
     Chest Beds Census Facil /
     selection=stepwise sle=0.20 sls=0.05;
run ;
```

Edited SAS output

(Complete output)

```
   Stepwise Procedure for Dependent Var RISK
Step 1 Var CULTURE Entrd R-sq=0.313 C(p)=47.48
           DF  Sum Sq  Mean Sq     F  Prob>F
Regression  1  62.9631 62.9631 50.49 0.0001
Error     111 138.4167  1.2470
Total     112 201.3798
           Par    Std   Type II
Variable   Est   Error   Sum Sq     F     Prob>F
INTERCEP 3.1979 0.1938 339.6491 272.37 0.0001
CULTURE  0.0733 0.0103  62.9631  50.49 0.0001
```

Edited SAS Output

```
Step 2 Var STAY Entered R-sq=0.450 C(p)=18.12
           DF  Sum Sqs Mean Sq   F   Prob>F
Regression  2  90.7020 45.3510 45.07 0.0001
Error     110 110.6778  1.00616214
Total     112 201.37982301
            Par   Std    Type II
Variable    Est   Error   Sum Sq     F Prob>F
INTERCEP  0.8055 0.4878  2.7440  2.73 0.1015
CULTURE   0.0565 0.0098 33.3969 33.19 0.0001
STAY      0.2755 0.0525 27.7388 27.57 0.0001
```

Edited SAS Output

```
Step 3 Var FACIL Entrd R-sq=0.493 C(p)=10.33
           DF Sum Sq   Mean Sq      F Prob>F
Regression  3  99.3608 33.1203 35.39 0.0001
Error     109 102.0190  0.9360
Total     112 201.3798
           Par   Std   Type II
Variable   Est  Error  Sum Sq     F  Prob>F
INTERCEP 0.4913 0.4816  0.9740  1.04 0.3099
CULTURE  0.0542 0.0095 30.5983 32.69 0.0001
STAY     0.2239 0.0534 16.4766 17.60 0.0001
FACIL    0.0196 0.0065  8.6588  9.25 0.0029
```

Edited SAS Output

```
Step 4 Var NRATIO Entered R-sq=0.525 C(p)=5.0278
                DF    Sum Sq   Mean Sq       F Prob>F
Regression       4     105.8210  26.4552   29.90 0.0001
Error          108      95.5589   0.88480418
Total          112     201.37982301
              Par        Std       Type II
Variable      Est         Error    Sum Sq     F Prob>F
INTERCEP   -0.4951   0.5938      0.6151    0.70 0.4063
CULTURE     0.0482   0.0095     22.8451   25.82 0.0001
STAY        0.2676   0.0543     21.4500   24.24 0.0001
NRATIO      0.7926   0.2933      6.4601    7.30 0.0080
FACIL       0.0175   0.0063      6.7535    7.63 0.0067
```

Edited SAS Output

```
Step 5 Var CHEST Entered R-sq=0.538 C(p)=4.19
           DF Sum Sq   Mean Sq     F  Prob>F
Regression  5 108.3272 21.6654 24.91 0.0001
Error     107  93.0527  0.8697
Total     112 201.3798
             Par   Std   Type II
Variable     Est  Error  Sum Sq     F  Prob>F
INTERCEP -0.7680 0.6102  1.3776  1.58 0.2109
CULTURE   0.0432 0.0098 16.7198 19.23 0.0001
STAY      0.2339 0.0574 14.4381 16.60 0.0001
NRATIO    0.6724 0.2993  4.3888  5.05 0.0267
CHEST     0.0092 0.0054  2.5062  2.88 0.0925
FACIL     0.0184 0.0063  7.4571  8.57 0.0042
-----------------------------------------
```

Edited SAS Output

```
Step 6 Var CHEST Removed R-sq=0.525 C(p)=5.03
             DF  Sum Sq   Mean Sq    F  Prob>F
Regression    4  105.8210 26.4552 29.90 0.0001
Error       108   95.5589  0.8848
Total       112  201.3799
            Par    Std   Type II
Variable    Est   Error  Sum Sq    F   Prob>F
INTERCEP -0.4951 0.5938  0.6151  0.70 0.4063
CULTURE   0.0482 0.0095 22.8451 25.82 0.0001
STAY      0.2676 0.0543 21.4500 24.24 0.0001
NRATIO    0.7926 0.2933  6.4601  7.30 0.0080
FACIL     0.0175 0.0063  6.7535  7.63 0.0067
-----------------------------------------
```

Edited SAS Output

All variables left in the model are significant at the 0.0500 level. The stepwise method terminated because the next variable to be entered was just removed.

Summary of Stepwise Proc for Dependent Var RISK

Step	Entd	Rem	In	R**2	Partl R**2	Model C(p)	F	Prob>F
1	CULTURE		1	0.313	0.313	47.48	50.49	0.0001
2	STAY		2	0.138	0.450	18.12	27.57	0.0001
3	FACIL		3	0.043	0.493	10.33	9.25	0.0029
4	NRATIO		4	0.032	0.526	5.03	7.30	0.0080
5	CHEST		5	0.012	0.538	4.19	2.88	0.0925
6		CHEST	4	0.012	0.526	5.02	2.88	0.0925

Comments on code and results

- Notice the option `selection` on the model lines.
- Forward adds variables until the smallest P-value is more than 0.5.
- Backward removes variables until all remaining are significant at 0.1 level.
- Final models for backward, forward and stepwise are virtually the same here: variables retained are Culture, Stay, Nratio, Chest and Facil. Exception is that forward retains Census at $P = 0.36$.
- Significance levels to add or delete variables are controlled by `sle=` and`sls=` options.

All Subsets

```
data scenic;
 infile 'scenic.dat' firstobs=2;
 input Stay  Age Risk Culture Chest Beds
     School Region Census Nurses Facil;
 Nratio = Nurses / Census  ;
proc reg  data=scenic;
  model Risk = Culture Stay Nurses Nratio
   Chest Beds Census Facil / selection=cp ;
run ;
```

Edited SAS output

(Complete output)

```
N = 113    Regression Models for Dependent Var: RISK
    C(p)   R-sq     Variables in Model
                     In
 4.19461 0.5379  5 CULTURE STAY NRATIO CHEST FACIL
 4.81202 0.5352  5 CULTURE STAY NRATIO CHEST CENSUS
 5.02783 0.5255  4 CULTURE STAY NRATIO FACIL
 5.33543 0.5329  5 CULTURE STAY NRATIO CHEST BEDS
 5.38786 0.5415  6 CULTURE STAY NRATIO CHEST CENSUS FACIL
 5.69350 0.5401  6 CULTURE STAY NRATIO CHEST BEDS FACIL
 5.89630 0.5392  6 CULTURE STAY NURSES NRATIO CHEST FACIL
 6.00546 0.5212  4 CULTURE STAY NRATIO CENSUS
 6.23202 0.5290  5 CULTURE STAY NURSES NRATIO CHEST
 6.47628 0.5191  4 CULTURE STAY NRATIO BEDS
 6.50213 0.5278  5 CULTURE STAY NRATIO CENSUS FACIL
 6.70444 0.5357  6 CULTURE STAY NURSES NRATIO CHEST CENSUS
 6.73959 0.5267  5 CULTURE STAY NRATIO BEDS FACIL
 6.77459 0.5354  6 CULTURE STAY NRATIO CHEST BEDS CENSUS
 6.91746 0.5260  5 CULTURE STAY NURSES NRATIO FACIL
```

```
 81.27048  0.1730   2  BEDS FACIL
 83.31964  0.1552   1  NURSES
 83.60929  0.1715   3  NURSES BEDS CENSUS
 84.59092  0.1584   2  NURSES CENSUS
 85.31844  0.1552   2  NURSES BEDS
 85.53858  0.1455   1  CENSUS
 86.28567  0.1510   2  BEDS CENSUS
 89.19019  0.1294   1  BEDS
111.09898  0.0332   1  NRATIO
```

Comments on code and results

- Every one of the 2^8 models was tried.
- Good possible models have C_p small and not too far from p because $E(C_p) = p$ when the model in question is correct.
- First listed model has C_6 a bit over 4.
- This is ok since only values larger than p can indicate a bias (resulting from a missing variable in the model.
- This method selects CULTURE, STAY, NRATIO, CHEST and FACIL as did BACKWARD and STEPWISE. (FORWARD also included CENSUS with a quite large P-value.)

Categorical Covariates and Variable Selection

- Conceptually it is easy to do variable selection in the same way when some of the variables are categorical.
- In SAS you have to use `proc reg` and that procedure has no facility for categorical variables.
- You create columns of the design matrix yourself and group together the columns which correspond to the categorical variable as follows:

SAS Code

```
data scenic;
 infile 'scenic.dat' firstobs=2;
 input Stay  Age Risk Culture Chest
  Beds School Region Census Nurses Facil;
 Nratio = Nurses / Census  ;
 R1 = -(Region-4)*(Region-3)*(Region-2)/6;
 R2 = (Region-4)*(Region-3)*(Region-1)/2;
 R3 = -(Region-4)*(Region-2)*(Region-1)/2;
 S1 = School-1;
proc reg  data=scenic;
  model Risk = S1 Culture Stay Nurses
    Nratio { R1 R2 R3 } Chest Beds Census
    Facil / selection=stepwise
    groupnames = 'School' 'Culture' 'Stay'
    'Nurses' 'Nratio' 'Region' 'Chest'
    'Beds' 'Census' 'Facil';
run ;
```

Comments on Code

- Variable R1 is 1 for cases in Region 1 and 0 for other cases. R2 is 1 for Region 2, R3 is 1 for Region 3.
- These 3 columns are the columns for the factor REGION using the corner point coding $\alpha_4 = 0$, in the notation of previous s.
- S1 is 1 for hospitals not attached to medical schools.
- Variables R1, R2 and R3 are grouped together by braces so that the selection method must put them all in or all out.
- The `data` step can be used more simply to compute R1, R2 and R3. See the HELP facility in SAS.
- `groupnames` names groups of variables so that, e.g., R1, R2 and R3 have a name, Region.

Edited SAS Output

(Complete output)

```
        Stepwise Procedure for Dependent Var RISK
Step 1 Group Culture Entered R-sq=0.313 C(p)=58.36
               Par  Std  Type II
Variable       Est  Error      Sum Sq      F  Prob>F
INTERCEP   3.19790 0.19377 339.64906 272.37 0.0001
--- Group Culture  ---      62.96314  50.49 0.0001
CULTURE    0.07326 0.01031  62.96314  50.49 0.0001
-----------------------------------------------------
Step 2 Group Stay Entered R-sq=0.450 C(p)=26.82
               Par   Std Type II
Variable       Est    Error    Sum Sq       F Prob>F
INTERCEP    0.8055 0.4878   2.7440   2.73 0.1015
--- Group Culture  ---      33.3969  33.19 0.0001
CULTURE     0.0565 0.0098   33.3969  33.19 0.0001
--- Group Stay     ---      27.7388  27.57 0.0001
STAY        0.2755 0.0525   27.7388  27.57 0.0001
```

```
Step 3 Group Facil Entered R-sq=0.493 C(p)=18.35
              Par    Std          Type II
Variable      Est    Error    Sum Sq       F Prob>F
INTERCEP    0.4913 0.4816   0.9740    1.04 0.3099
--- Group Culture  ---     30.5983   32.69 0.0001
CULTURE     0.0542 0.0095 30.5983   32.69 0.0001
--- Group Stay     ---     16.4766   17.60 0.0001
STAY        0.2239 0.0534 16.4766   17.60 0.0001
--- Group Facil    ---      8.6588    9.25 0.0029
FACIL       0.0196 0.0065  8.6588    9.25 0.0029
-----------------------------------------------------
```

```
Step 4 Group Nratio Entered R-sq=0.526 C(p)=12.54
              Par    Std          Type II
Variable      Est    Error   Sum Sq       F Prob>F
INTERCEP   -0.4951 0.593    0.6151    0.70 0.4063
--- Group Culture  ---     22.8451   25.82 0.0001
CULTURE     0.0482 0.0095 22.8451   25.82 0.0001
--- Group Stay     ---     21.4500   24.24 0.0001
STAY        0.2676 0.0543 21.4500   24.24 0.0001
--- Group Nratio   ---      6.4601    7.30 0.0080
NRATIO      0.7926 0.2933  6.4601    7.30 0.0080
--- Group Facil    ---      6.7535    7.63 0.0067
FACIL       0.0175 0.0063  6.7535    7.63 0.0067
-----------------------------------------------
```

```
Step 5 Group Chest Entered R-sq=0.538 C(p)=11.51
             Par    Std          Type II
Variable     Est    Error   Sum Sq     F   Prob>F
INTERCEP   -0.7680 0.6102   1.3776 1.58   0.2109
--- Group Culture  ---     16.7198    19.23  0.0001
CULTURE      0.0432 0.0098 16.7198    19.23  0.0001
--- Group Stay     ---     14.4381    16.60  0.0001
STAY         0.2339 0.0574 14.4381    16.60  0.0001
--- Group Nratio   ---      4.3888     5.05  0.0267
NRATIO       0.6724 0.2993  4.3888     5.05  0.0267
--- Group Chest    ---      2.5062     2.88  0.0925
CHEST        0.0092 0.0054  2.5062     2.88  0.0925
--- Group Facil    ---      7.4571     8.57  0.0042
FACIL        0.0184 0.0063  7.4571     8.57  0.0042
-----------------------------------------------------
```

Step 6 Group Region Entered R-sq=0.568 C(p)=10.13

Variable	Par Est	Std Error	Type II Sum Sq	F	Prob>F
INTERCEP	-0.6616	0.6893	0.7700	0.92	0.3394
--- Group Culture		---	19.4185	23.23	0.0001
CULTURE	0.0472	0.0098	19.4185	23.23	0.0001
--- Group Stay		---	18.6472	22.31	0.0001
STAY	0.2841	0.0602	18.6472	22.31	0.0001
--- Group Nratio		---	1.8677	2.23	0.1380
NRATIO	0.4774	0.3194	1.8677	2.23	0.1380
--- Group Region		---	6.1086	2.44	0.0689
R1	-0.9115	0.3383	6.0688	7.26	0.0082
R2	-0.6117	0.3063	3.3341	3.99	0.0484
R3	-0.5401	0.3053	2.6157	3.13	0.0799
--- Group Chest		---	3.1059	3.72	0.0566
CHEST	0.0103	0.0053	3.1059	3.72	0.0566
--- Group Facil		---	7.6625	9.17	0.0031
FACIL	0.0188	0.0062	7.6625	9.17	0.0031

```
Step 7 Group School Entered R-sq=0.578 C(p)=9.68
              Par    Std        Type II
Variable      Est    Error   Sum Sq      F Prob>F
INTERCEP   -1.2931 0.7944   2.1845   2.65 0.1066
--- Group School   ---      2.0234   2.45 0.1203
S1          0.4587 0.2928   2.0234   2.45 0.1203
--- Group Culture  ---     21.1424 25.64 0.0001
CULTURE     0.0502 0.0099 21.1424 25.64 0.0001
--- Group Stay     ---     19.9084 24.15 0.0001
STAY        0.2958 0.0602 19.9084 24.15 0.0001
--- Group Nratio   ---      1.4288   1.73 0.1909
NRATIO      0.4203 0.3192   1.4288   1.73 0.1909
--- Group Region   ---      7.0904   2.87 0.0402
R1         -0.9974 0.3404   7.0775   8.58 0.0042
R2         -0.6443 0.3049   3.6812   4.46 0.0370
R3         -0.5995 0.3056   3.1735   3.85 0.0525
--- Group Chest    ---      2.8545   3.46 0.0656
CHEST       0.0099 0.0053   2.8545   3.46 0.0656
--- Group Facil    ---      9.6853 11.75 0.0009
```

```
Step 8 Group Nratio Removed R-sq=0.571 C(p)=9.41
              Par    Std          Type II
Variable      Est    Error   Sum Sq      F Prob>F
INTERCEP   -0.8324 0.7157   1.1231   1.35 0.2475
--- Group School   ---      2.4623   2.97 0.0880
S1          0.5027 0.2919   2.4623   2.97 0.0880
--- Group Culture  ---     23.6669 28.50 0.0001
CULTURE     0.0523 0.0098 23.6669 28.50 0.0001
--- Group Stay     ---     18.4796 22.26 0.0001
STAY        0.2747 0.0582 18.4796 22.26 0.0001
--- Group Region   ---      9.6872   3.89 0.0111
R1         -1.1070 0.3312   9.2728 11.17 0.0012
R2         -0.7667 0.2914   5.7492   6.92 0.0098
R3         -0.7594 0.2814   6.0465   7.28 0.0081
--- Group Chest    ---      3.9212   4.72 0.0320
CHEST       0.0113 0.0052   3.9212   4.72 0.0320
--- Group Facil    ---     11.3028 13.61 0.0004
FACIL       0.0255 0.0069 11.3028 13.61 0.0004
```

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

Summary of Stepwise Proc for Dependent Var RISK

	Group	Num	Partl	Model			
Step	Entd Remd	In	R**2	R**2	C(p)	F	Prob>F
1	Culture	1	0.312	0.313	58.36	50.49	0.0000
2	Stay	2	0.138	0.450	26.82	27.57	0.0000
3	Facil	3	0.043	0.493	18.35	9.25	0.0029
4	Nratio	4	0.032	0.526	12.54	7.30	0.0080
5	Chest	5	0.012	0.538	11.51	2.88	0.0925
6	Region	8	0.030	0.568	10.13	2.44	0.0689
7	School	9	0.010	0.578	9.68	2.45	0.1203
8	Nratio	8	0.007	0.571	9.41	1.73	0.1909

Comments on Output

- Final model selected has variables SCHOOL, CULTURE, STAY, REGION, CHEST and FACIL.
- Variable NRATIO included at step 4 was eliminated at step 8.
- `groupnames` assigns names to groups of variables.
- So the printout is more easily read.