STAT 350: Lecture 27

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.

Thus far we have used F and t tests to compare 2 models at a time. We 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. Here we investigate 4 mechanical (more or less) variable selection methods: Forward, Backward, Stepwise and All Subsets.

FORWARD

- 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

- 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

- Start with model with no predictors.
- Add variable with largest F-statistic (provided P more 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 the 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 and the class notes for that .)

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 ;
```

EDITED SAS OUTPUT

```
    Forward Selection Procedure for Dependent Variable RISK
Step 1 Variable CULTURE Entered R-square=0.312659 C(p)=47.4779
            DF Sum of Squares   Mean Square     F Prob>F
Regression   1   62.96314170   62.96314170  50.49 0.0001
Error      111  138.41668131    1.24699713
Total      112  201.37982301
           Parameter   Standard       Type II
Variable    Estimate      Error Sum of Squares     F Prob>F
INTERCEP  3.19789965 0.19376813  339.64905575 272.37 0.0001
CULTURE   0.07325862 0.01030975   62.96314170  50.49 0.0001
----------------------------------------------------------
Step 2 Variable STAY Entered R-square=0.45040 C(p)=18.1196
             DF Sum of Squares Mean Square   F   Prob>F
Regression    2  90.70198757   45.35099379 45.07 0.0001
Error       110 110.67783543    1.0061621
Total       112 201.37982301
           Parameter   Standard      Type II
Variable    Estimate      Error Sum of Squares     F Prob>F
INTERCEP  0.80549102 0.48775579     2.74400250  2.73 0.1015
CULTURE   0.05645147 0.00979843    33.39687778 33.19 0.0001
STAY      0.27547211 0.05246473    27.73884588 27.57 0.0001
----------------------------------------------------------
Step 3 Variable FACIL Entered R-square=0.4934 C(p)=10.3309
            DF Sum of Squares      Mean Square   F   Prob>F
Regression   3   99.36082444      33.12027481  35.39 0.0001
Error      109  102.01899857       0.93595412
```

```
Total      112  201.37982301
           Parameter   Standard       Type II
Variable    Estimate      Error Sum of Squares   F Prob>F
INTERCEP  0.49133226 0.48163614    0.97401801  1.04 0.3099
CULTURE   0.05419997 0.00947933   30.59827862 32.69 0.0001
STAY      0.22390748 0.05336561   16.47664606 17.60 0.0001
FACIL     0.01963027 0.00645392    8.65883687  9.25 0.0029
-------------------------------------------------------
Step 4 Variable NRATIO Entered R-square=0.52548 C(p)= 5.0278
             DF Sum of Squares Mean Square    F   Prob>F
Regression    4 105.82097194    26.45524298 29.90 0.0001
Error       108  95.55885107     0.88480418
Total       112 201.37982301
             Parameter   Standard       Type II
Variable      Estimate      Error Sum of Squares   F Prob>F
INTERCEP   -0.49505513 0.59376426    0.61507231  0.70 0.4063
CULTURE     0.04818092 0.00948204   22.84513509 25.82 0.0001
STAY        0.26758404 0.05434637   21.44995791 24.24 0.0001
NRATIO      0.79262357 0.29333869    6.46014750  7.30 0.0080
FACIL       0.01747585 0.00632554    6.75349077  7.63 0.0067
-----------------------------------------------------------
Step 5 Variable CHEST Entered R-square=0.53792 C(p)= 4.19461
             DF   Sum of Squares  Mean Square  F  Prob>F
Regression    5   108.32716704  21.66543341  24.91 0.0001
Error       107    93.05265597   0.86965099
Total       112   201.37982301
           Parameter   Standard        Type II
Variable    Estimate      Error Sum of Squares    F Prob>F
INTERCEP -0.76804342 0.61022741   1.37763165   1.58 0.2109
CULTURE   0.04318856 0.00984976  16.71979631  19.23 0.0001
STAY      0.23392650 0.05741114  14.43814950  16.60 0.0001
NRATIO    0.67240318 0.29931440   4.38883521   5.05 0.0267
CHEST     0.00917860 0.00540681   2.50619510   2.88 0.0925
FACIL     0.01843860 0.00629673   7.45710068   8.57 0.0042
-----------------------------------------------------------
Step 6 Variable CENSUS Entered R-square=0.54147 C(p)=5.38786
            DF Sum of Squares  Mean Square     F Prob>F
Regression   6   109.04079737  18.17346623 20.86 0.0001
```

```
Error      106    92.33902564   0.87112288
Total      112   201.37982301
            Parameter    Standard     Type II
Variable     Estimate       Error Sum of Squares   F Prob>F
INTERCEP  -0.60982957  0.63526660   0.80275743  0.92 0.3393
CULTURE    0.04327612  0.00985857  16.78604192 19.27 0.0001
STAY       0.21806849  0.06007156  11.47961755 13.18 0.0004
NRATIO     0.74253899  0.30942748   5.01649265  5.76 0.0182
CHEST      0.00967176  0.00543875   2.75481205  3.16 0.0782
CENSUS     0.00092337  0.00102018   0.71363033  0.82 0.3675
FACIL      0.01171496  0.00974167   1.25977812  1.45 0.2318
------------------------------------------------------------
No other variable met the 0.5000 significance level
for entry into the model.  Summary of Forward Selection
Procedure for Dependent Variable RISK
     Variable Number Partial  Model
Step  Entered  In      R**2     R**2     C(p)        F Prob>F
 1    CULTURE   1    0.3127   0.3127  47.4779 50.4918 0.0001
 2    STAY      2    0.1377   0.4504  18.1196 27.5690 0.0001
 3    FACIL     3    0.0430   0.4934  10.3309  9.2513 0.0029
 4    NRATIO    4    0.0321   0.5255   5.0278  7.3012 0.0080
 5    CHEST     5    0.0124   0.5379   4.1946  2.8818 0.0925
 6    CENSUS    6    0.0035   0.5415   5.3879  0.8192 0.3675
```

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 ;
```

EDITED SAS OUTPUT

```
 Backward Elimination Procedure for Dependent Variable RISK
Step 0 All Variables Entered R-square=0.543172 C(p)=9.00000
             DF Sum of Squares Mean Square     F Prob>F
Regression    8   109.38389081 13.67298635 15.46 0.0001
Error       104    91.99593220  0.88457627
Total       112   201.37982301
          Parameter   Standard        Type II
Variable   Estimate      Error Sum of Squares    F  Prob>F
INTERCEP -0.61544324 0.66644759     0.75436112  0.85 0.3579
CULTURE   0.04410334 0.01004539    17.05081019 19.28 0.0001
STAY      0.20541554 0.06405141     9.09797087 10.29 0.0018
NURSES   -0.00087592 0.00216138     0.14527948  0.16 0.6861
NRATIO    0.85012850 0.39334446     4.13198139  4.67 0.0330
CHEST     0.00946882 0.00549665     2.62501031  2.97 0.0879
BEDS     -0.00106503 0.00265225     0.14263677  0.16 0.6888
CENSUS    0.00295314 0.00357645     0.60311345  0.68 0.4109
FACIL     0.01312502 0.01010817     1.49138778  1.69 0.1970
-----------------------------------------------------------
Step 1 Variable BEDS Removed R-square=0.54246375 C(p)=7.161249
             DF Sum of Squares Mean Square     F  Prob>F
Regression    7  109.24125404  15.6058934  17.78 0.0001
Error       105   92.13856897   0.87751018
Total       112  201.37982301
             Parameter   Standard        Type II
Variable      Estimate      Error Sum of Squares    F Prob>F
INTERCEP   -0.66993458 0.64987468    0.93251908  1.06 0.3050
CULTURE     0.04396605 0.00999939   16.96447009 19.33 0.0001
STAY        0.21222554 0.06151830   10.44332482 11.90 0.0008
NURSES     -0.00101551 0.00212470    0.20045667  0.23 0.6337
NRATIO      0.85643956 0.39145742    4.20026358  4.79 0.0309
CHEST       0.00947189 0.00547465    2.62672060  2.99 0.0865
CENSUS      0.00178564 0.00207440    0.65020801  0.74 0.3913
FACIL       0.01228514 0.00984983    1.36506916  1.56 0.2151
---------------------------------------------------------
Step 2 Variable NURSES Removed R-square=0.54147 C(p)=5.3879
             DF   Sum of Squares   Mean Square   F   Prob>F
Regression    6     109.04079737   18.17346623 20.86 0.0001
Error       106      92.33902564    0.87112288
```

```
Total       112     201.37982301
             Parameter   Standard        Type II
Variable      Estimate      Error Sum of Squares   F Prob>F
INTERCEP   -0.60982957 0.63526660   0.80275743  0.92 0.3393
CULTURE     0.04327612 0.00985857  16.78604192 19.27 0.0001
STAY        0.21806849 0.06007156  11.47961755 13.18 0.0004
NRATIO      0.74253899 0.30942748   5.01649265  5.76 0.0182
CHEST       0.00967176 0.00543875   2.75481205  3.16 0.0782
CENSUS      0.00092337 0.00102018   0.71363033  0.82 0.3675
FACIL       0.01171496 0.00974167   1.25977812  1.45 0.2318
----------------------------------------------------------
Step 3 Variable CENSUS Removed R-square=0.5379 C(p)=4.1946
             DF Sum of Squares   Mean Square    F  Prob>F
Regression    5  108.32716704   21.66543341  24.91 0.0001
Error       107   93.05265597    0.86965099
Total       112  201.37982301
           Parameter   Standard        Type II
Variable    Estimate      Error Sum of Squares    F Prob>F
INTERCEP -0.76804342 0.61022741    1.37763165  1.58 0.2109
CULTURE   0.04318856 0.00984976   16.71979631 19.23 0.0001
STAY      0.23392650 0.05741114   14.43814950 16.60 0.0001
NRATIO    0.67240318 0.29931440    4.38883521  5.05 0.0267
CHEST     0.00917860 0.00540681    2.50619510  2.88 0.0925
FACIL     0.01843860 0.00629673    7.45710068  8.57 0.0042
--------------------------------------------------------
All variables left in the model are significant at the
0.1000 level.  Summary of Backward Elimination Procedure
for Dependent Variable RISK
     Variable  Number Partial  Model
Step Removed    In     R**2     R**2   C(p)      F Prob>F
 1   BEDS       7    0.0007   0.5425 7.1612 0.1612 0.6888
 2   NURSES     6    0.0010   0.5415 5.3879 0.2284 0.6337
 3   CENSUS     5    0.0035   0.5379 4.1946 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

```
               Stepwise Procedure for Dependent Variable RISK
Step 1 Variable CULTURE Entered R-square=0.312659 C(p)=47.47795
            DF       Sum of Squares   Mean Square      F Prob>F
Regression  1          62.96314170   62.96314170   50.49 0.0001
Error     111         138.41668131    1.24699713
Total     112         201.37982301
           Parameter      Standard       Type II
Variable    Estimate         Error Sum of Squares      F Prob>F
INTERCEP  3.19789965    0.19376813   339.64905575 272.37 0.0001
CULTURE   0.07325862    0.01030975    62.96314170  50.49 0.0001
---------------------------------------------------------------
Step 2 Variable STAY Entered R-square=0.45040 C(p)=18.119607
                DF    Sum of Squares   Mean Square     F Prob>F
Regression       2       90.70198757   45.35099379 45.07 0.0001
Error          110      110.67783543    1.00616214
Total          112      201.37982301
              Parameter  Standard         Type II
Variable       Estimate     Error  Sum of Squares     F Prob>F
INTERCEP     0.80549102  0.48775579    2.74400250  2.73 0.1015
CULTURE      0.05645147  0.00979843   33.39687778 33.19 0.0001
STAY         0.27547211  0.05246473   27.73884588 27.57 0.0001
-------------------------------------------------------------
Step 3 Variable FACIL Entered R-square=0.49340 C(p)=10.33092
                DF   Sum of Squares  Mean Square     F Prob>F
Regression       3      99.36082444  33.12027481 35.39 0.0001
Error          109     102.01899857   0.93595412
Total          112     201.37982301
            Parameter     Standard        Type II
Variable     Estimate        Error Sum of Squares     F Prob>F
```

```
INTERCEP   0.49133226   0.48163614     0.97401801  1.04 0.3099
CULTURE    0.05419997   0.00947933    30.59827862 32.69 0.0001
STAY       0.22390748   0.05336561    16.47664606 17.60 0.0001
FACIL      0.01963027   0.00645392     8.65883687  9.25 0.0029
-------------------------------------------------------
Step 4 Variable NRATIO Entered R-square=0.5254795 C(p)=5.0278
              DF   Sum of Squares  Mean Square      F Prob>F
Regression     4     105.82097194  26.45524298  29.90 0.0001
Error        108      95.55885107   0.88480418
Total        112     201.37982301
            Parameter      Standard      Type II
Variable     Estimate         Error   Sum of Squares    F Prob>F
INTERCEP  -0.49505513  0.59376426     0.61507231   0.70 0.4063
CULTURE    0.04818092  0.00948204    22.84513509  25.82 0.0001
STAY       0.26758404  0.05434637    21.44995791  24.24 0.0001
NRATIO     0.79262357  0.29333869     6.46014750   7.30 0.0080
FACIL      0.01747585  0.00632554     6.75349077   7.63 0.0067
-------------------------------------------------------------
Step 5 Variable CHEST Entered R-square=0.53792 C(p)=4.19461
                DF   Sum of Squares  Mean Square     F Prob>F
Regression       5     108.32716704  21.66543341 24.91 0.0001
Error          107      93.05265597   0.86965099
Total          112     201.37982301
            Parameter    Standard        Type II
Variable     Estimate         Error Sum of Squares       F Prob>F
INTERCEP  -0.76804342  0.61022741     1.37763165   1.58 0.2109
CULTURE    0.04318856  0.00984976    16.71979631  19.23 0.0001
STAY       0.23392650  0.05741114    14.43814950  16.60 0.0001
NRATIO     0.67240318  0.29931440     4.38883521   5.05 0.0267
CHEST      0.00917860  0.00540681     2.50619510   2.88 0.0925
FACIL      0.01843860  0.00629673     7.45710068   8.57 0.0042
-------------------------------------------------------------
Step 6 Variable CHEST Removed R-square=0.5254795 C(p)=5.027826
            DF      Sum of Squares   Mean Square      F Prob>F
Regression   4        105.82097194   26.45524298  29.90 0.0001
Error      108         95.55885107    0.88480418
Total      112        201.37982301
            Parameter     Standard        Type II
```

```
Variable     Estimate        Error Sum of Squares      F  Prob>F
INTERCEP  -0.49505513   0.59376426    0.61507231    0.70 0.4063
CULTURE    0.04818092   0.00948204   22.84513509   25.82 0.0001
STAY       0.26758404   0.05434637   21.44995791   24.24 0.0001
NRATIO     0.79262357   0.29333869    6.46014750    7.30 0.0080
FACIL      0.01747585   0.00632554    6.75349077    7.63 0.0067
----------------------------------------------------------------
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 Procedure for Dependent Variable RISK
     Variable       Number Partial  Model
Step Entered Remove   In     R**2    R**2    C(p)        F Prob>F
 1   CULTURE           1  0.3127  0.3127 47.4779 50.4918 0.0001
 2   STAY              2  0.1377  0.4504 18.1196 27.5690 0.0001
 3   FACIL             3  0.0430  0.4934 10.3309  9.2513 0.0029
 4   NRATIO            4  0.0321  0.5255  5.0278  7.3012 0.0080
 5   CHEST             5  0.0124  0.5379  4.1946  2.8818 0.0925
 6           CHEST     4  0.0124  0.5255  5.0278  2.8818 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 tthat 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

```
N = 113    Regression Models for Dependent Variable: RISK
    C(p)   R-square     Variables in Model
                     In
 4.19461 0.53792463  5 CULTURE STAY NRATIO CHEST FACIL
 4.81202 0.53521260  5 CULTURE STAY NRATIO CHEST CENSUS
 5.02783 0.52547952  4 CULTURE STAY NRATIO FACIL
 5.33543 0.53291351  5 CULTURE STAY NRATIO CHEST BEDS
 5.38786 0.54146833  6 CULTURE STAY NRATIO CHEST CENSUS FACIL
 5.69350 0.54012581  6 CULTURE STAY NRATIO CHEST BEDS FACIL
 5.89630 0.53923499  6 CULTURE STAY NURSES NRATIO CHEST FACIL
 6.00546 0.52118519  4 CULTURE STAY NRATIO CENSUS
 6.23202 0.52897517  5 CULTURE STAY NURSES NRATIO CHEST
 6.47628 0.51911707  4 CULTURE STAY NRATIO BEDS
 6.50213 0.52778865  5 CULTURE STAY NRATIO CENSUS FACIL
 6.70444 0.53568517  6 CULTURE STAY NURSES NRATIO CHEST CENSUS
 6.73959 0.52674562  5 CULTURE STAY NRATIO BEDS FACIL
 6.77459 0.53537702  6 CULTURE STAY NRATIO CHEST BEDS CENSUS
 6.91746 0.52596429  5 CULTURE STAY NURSES NRATIO FACIL
```

⋮

```
  81.27048  0.17300751   2  BEDS FACIL
  83.31964  0.15522130   1  NURSES
  83.60929  0.17151925   3  NURSES BEDS CENSUS
  84.59092  0.15842223   2  NURSES CENSUS
  85.31844  0.15522654   2  NURSES BEDS
  85.53858  0.14547441   1  CENSUS
  86.28567  0.15097790   2  BEDS CENSUS
  89.19019  0.12943445   1  BEDS
 111.09898  0.03319840   1  NRATIO
```

Comments on code and results

- Every one of the 2^{11} 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.

INCLUDING CATEGORICAL COVARIATES

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:

```
options pagesize=60 linesize=80;
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

```
              Stepwise Procedure for Dependent Variable RISK
Step 1 Group Culture Entered R-square=0.3126586 C(p)=58.364132
             Parameter   Standard        Type II
Variable      Estimate      Error Sum of Squares      F  Prob>F
INTERCEP    3.19789965 0.19376813   339.64905575 272.37 0.0001
--- Group Culture  ---               62.96314170  50.49 0.0001
CULTURE     0.07325862 0.01030975    62.96314170  50.49 0.0001
------------------------------------------------------------
Step 2 Group Stay Entered R-square=0.45040256 C(p)=26.82418731
             Parameter   Standard        Type II
Variable      Estimate      Error Sum of Squares       F Prob>F
INTERCEP    0.80549102 0.48775579     2.74400250    2.73 0.1015
--- Group Culture  ---               33.39687778   33.19 0.0001
CULTURE     0.05645147 0.00979843    33.39687778   33.19 0.0001
--- Group Stay     ---               27.73884588   27.57 0.0001
STAY        0.27547211 0.05246473    27.73884588   27.57 0.0001
------------------------------------------------------------
Step 3 Group Facil Entered R-square=0.49340010 C(p)=18.3545047
             Parameter   Standard        Type II
Variable      Estimate      Error Sum of Squares       F Prob>F
INTERCEP    0.49133226 0.48163614     0.97401801    1.04 0.3099
--- Group Culture  ---               30.59827862   32.69 0.0001
```

```
CULTURE     0.05419997 0.00947933    30.59827862  32.69 0.0001
--- Group Stay     ---               16.47664606  17.60 0.0001
STAY        0.22390748 0.05336561    16.47664606  17.60 0.0001
--- Group Facil    ---                8.65883687   9.25 0.0029
FACIL       0.01963027 0.00645392     8.65883687   9.25 0.0029
------------------------------------------------------------
Step 4 Group Nratio Entered R-square=0.52547952 C(p)=12.543329
             Parameter   Standard        Type II
Variable      Estimate      Error Sum of Squares      F Prob>F
INTERCEP   -0.49505513 0.59376426     0.61507231   0.70 0.4063
--- Group Culture  ---               22.84513509  25.82 0.0001
CULTURE     0.04818092 0.00948204    22.84513509  25.82 0.0001
--- Group Stay     ---               21.44995791  24.24 0.0001
STAY        0.26758404 0.05434637    21.44995791  24.24 0.0001
--- Group Nratio   ---                6.46014750   7.30 0.0080
NRATIO      0.79262357 0.29333869     6.46014750   7.30 0.0080
--- Group Facil    ---                6.75349077   7.63 0.0067
FACIL       0.01747585 0.00632554     6.75349077   7.63 0.0067
--------------------------------------------------------
Step 5 Group Chest Entered R-square=0.53792463 C(p)=11.513007
             Parameter   Standard        Type II
Variable      Estimate      Error Sum of Squares    F  Prob>F
INTERCEP   -0.76804342 0.61022741     1.37763165 1.58  0.2109
--- Group Culture  ---            16.71979631   19.23  0.0001
CULTURE     0.04318856 0.00984976 16.71979631   19.23  0.0001
--- Group Stay     ---            14.43814950   16.60  0.0001
STAY        0.23392650 0.05741114 14.43814950   16.60  0.0001
--- Group Nratio   ---             4.38883521    5.05  0.0267
NRATIO      0.67240318 0.29931440  4.38883521    5.05  0.0267
--- Group Chest    ---             2.50619510    2.88  0.0925
CHEST       0.00917860 0.00540681  2.50619510    2.88  0.0925
--- Group Facil    ---             7.45710068    8.57  0.0042
FACIL       0.01843860 0.00629673  7.45710068    8.57  0.0042
-----------------------------------------------------------
Step 6 Group Region Entered R-square=0.5682584 C(p)=10.126881
             Parameter   Standard        Type II
Variable      Estimate      Error Sum of Squares     F Prob>F
INTERCEP   -0.66156855 0.68931767     0.77004723  0.92 0.3394
```

```
--- Group Culture  ---                19.41848300 23.23 0.0001
CULTURE     0.04717749 0.00978882    19.41848300 23.23 0.0001
--- Group Stay     ---                18.64724032 22.31 0.0001
STAY        0.28408192 0.06015054    18.64724032 22.31 0.0001
--- Group Nratio   ---                 1.86769604  2.23 0.1380
NRATIO      0.47735146 0.31936579     1.86769604  2.23 0.1380
--- Group Region   ---                 6.10861501  2.44 0.0689
R1         -0.91152625 0.33831556     6.06877293  7.26 0.0082
R2         -0.61170886 0.30630883     3.33408744  3.99 0.0484
R3         -0.54005754 0.30531855     2.61565335  3.13 0.0799
--- Group Chest    ---                 3.10587423  3.72 0.0566
CHEST       0.01029102 0.00533912     3.10587423  3.72 0.0566
--- Group Facil    ---                 7.66252029  9.17 0.0031
FACIL       0.01883340 0.00622080     7.66252029  9.17 0.0031
-----------------------------------------------------------
Step 7 Group School Entered R-square=0.57830628 C(p)=9.680280
             Parameter   Standard       Type II
Variable      Estimate      Error Sum of Squares     F Prob>F
INTERCEP   -1.29313397 0.79443852     2.18445103  2.65 0.1066
--- Group School   ---                 2.02343484  2.45 0.1203
S1          0.45874175 0.29282732     2.02343484  2.45 0.1203
--- Group Culture  ---                21.14238169 25.64 0.0001
CULTURE     0.05016596 0.00990650    21.14238169 25.64 0.0001
--- Group Stay     ---                19.90843811 24.15 0.0001
STAY        0.29583936 0.06020399    19.90843811 24.15 0.0001
--- Group Nratio   ---                 1.42881407  1.73 0.1909
NRATIO      0.42026288 0.31924279     1.42881407  1.73 0.1909
--- Group Region   ---                 7.09035688  2.87 0.0402
R1         -0.99737538 0.34041455     7.07745167  8.58 0.0042
R2         -0.64425716 0.30489819     3.68115979  4.46 0.0370
R3         -0.59950685 0.30557155     3.17349874  3.85 0.0525
--- Group Chest    ---                 2.85453005  3.46 0.0656
CHEST       0.00987802 0.00530873     2.85453005  3.46 0.0656
--- Group Facil    ---                 9.68526975 11.75 0.0009
FACIL       0.02391008 0.00697611     9.68526975 11.75 0.0009
------------------------------------------------------------
Step 8 Group Nratio Removed R-square=0.57121116 C(p)=9.407905
             Parameter   Standard        Type II
```

```
Variable      Estimate      Error Sum of Squares     F Prob>F
INTERCEP   -0.83240584 0.71570292     1.12313185  1.35 0.2475
--- Group School   ---                2.46231681  2.97 0.0880
S1          0.50274483 0.29193670     2.46231681  2.97 0.0880
--- Group Culture  ---               23.66688888 28.50 0.0001
CULTURE     0.05233635 0.00980270    23.66688888 28.50 0.0001
--- Group Stay     ---               18.47964968 22.26 0.0001
STAY        0.27469386 0.05822575    18.47964968 22.26 0.0001
--- Group Region   ---                9.68716458  3.89 0.0111
R1         -1.10696516 0.33123989     9.27275385 11.17 0.0012
R2         -0.76673818 0.29137725     5.74922078  6.92 0.0098
R3         -0.75936643 0.28139304     6.04647398  7.28 0.0081
--- Group Chest    ---                3.92124933  4.72 0.0320
CHEST       0.01132621 0.00521177     3.92124933  4.72 0.0320
--- Group Facil    ---               11.30278424 13.61 0.0004
FACIL       0.02545939 0.00690031    11.30278424 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 the model.
     Summary of Stepwise Procedure for Dependent Variable RISK
      Group         Number Partial Model
Step Entered Removed In    R**2    R**2    C(p)       F Prob>F
   1 Culture         1  0.3127  0.3127 58.3641 50.4918 0.0000
   2 Stay            2  0.1377  0.4504 26.8242 27.5690 0.0000
   3 Facil           3  0.0430  0.4934 18.3545  9.2513 0.0029
   4 Nratio          4  0.0321  0.5255 12.5433  7.3012 0.0080
   5 Chest           5  0.0124  0.5379 11.5130  2.8818 0.0925
   6 Region          8  0.0303  0.5683 10.1269  2.4357 0.0689
   7 School          9  0.0100  0.5783  9.6803  2.4542 0.1203
   8         Nratio  8  0.0071  0.5712  9.4079  1.7330 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.

- It is possible to name the groups so that the printout is more easily read.