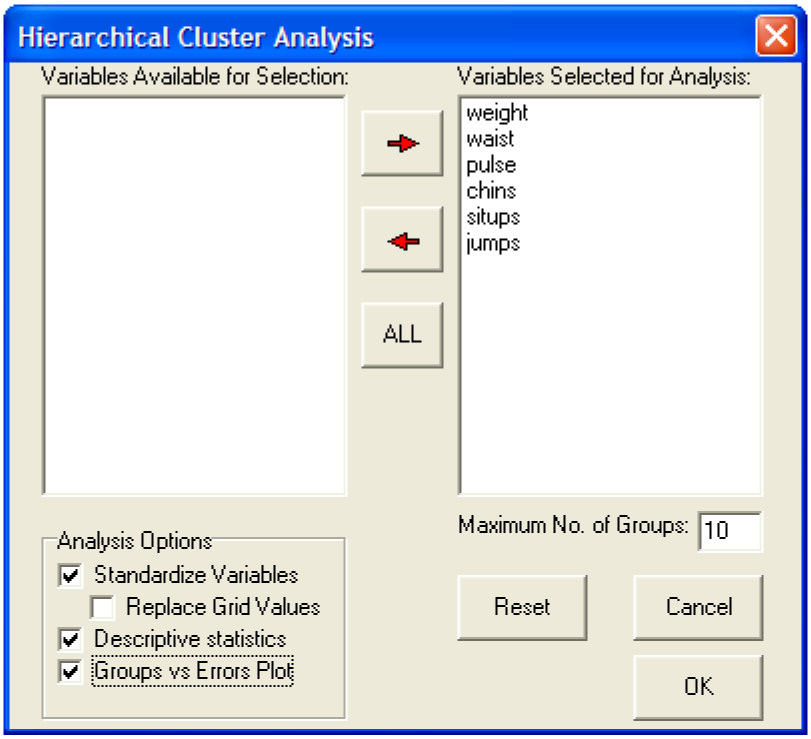
Hierarchical Cluster Analyses

To demonstrate the Hierarchical Clustering program, the data to be analyzed is the one labeled cansas.LAZ. You will see the form below with specifications for the grouping:



Specifications fo the Hierarchical Cluster Analysis

Results for the hierarchical analysis that you would obtain after clicking the Compute button are presented below:

```
Hierarchical Cluster Analysis

Number of object to cluster = 20 on 6 variables.

Variable Means
Variables      weight     waist      pulse      chins     situps      jumps
              178.600    35.400     56.100      9.450    145.550     70.300


Variable Variances
Variables      weight     waist      pulse      chins     situps      jumps
              609.621    10.253     51.989     27.945   3914.576   2629.379


Variable Standard Deviations
Variables      weight     waist      pulse      chins     situps      jumps
               24.691     3.202      7.210      5.286     62.567     51.277

19 groups after combining group 1 (n = 1 ) and group 5 (n = 1) error =   0.386
18 groups after combining group 17 (n = 1 ) and group 18 (n = 1) error =   0.387
17 groups after combining group 11 (n = 1 ) and group 17 (n = 2) error =   0.556
16 groups after combining group 1 (n = 2 ) and group 16 (n = 1) error =   0.663
15 groups after combining group 3 (n = 1 ) and group 7 (n = 1) error =   0.805
14 groups after combining group 4 (n = 1 ) and group 10 (n = 1) error =   1.050
```

```
13 groups after combining group 2 (n = 1 ) and group 6 (n = 1) error =   1.345
12 groups after combining group 1 (n = 3 ) and group 14 (n = 1) error =   1.402
11 groups after combining group 0 (n = 1 ) and group 1 (n = 4) error =   1.489
10 groups after combining group 11 (n = 3 ) and group 12 (n = 1) error =   2.128
Group 1 (n= 5)
   Object = CASE 1
   Object = CASE 2
   Object = CASE 6
   Object = CASE 15
   Object = CASE 17
Group 3 (n= 2)
   Object = CASE 3
   Object = CASE 7
Group 4 (n= 2)
   Object = CASE 4
   Object = CASE 8
Group 5 (n= 2)
   Object = CASE 5
   Object = CASE 11
Group 9 (n= 1)
   Object = CASE 9
Group 10 (n= 1)
   Object = CASE 10
Group 12 (n= 4)
   Object = CASE 12
   Object = CASE 13
   Object = CASE 18
   Object = CASE 19
Group 14 (n= 1)
   Object = CASE 14
Group 16 (n= 1)
   Object = CASE 16
Group 20 (n= 1)
   Object = CASE 20

(…. for 9 groups, 8 groups, etc. down to 2 groups)

4 groups after combining group 4 (n = 6 ) and group 9 (n = 1) error =  11.027
Group 1 (n= 8)
   Object = CASE 1
   Object = CASE 2
   Object = CASE 3
   Object = CASE 6
   Object = CASE 7
   Object = CASE 15
   Object = CASE 16
   Object = CASE 17
Group 4 (n= 4)
   Object = CASE 4
   Object = CASE 8
   Object = CASE 9
   Object = CASE 20
Group 5 (n= 7)
   Object = CASE 5
   Object = CASE 10
   Object = CASE 11
   Object = CASE 12
   Object = CASE 13
   Object = CASE 18
   Object = CASE 19
Group 14 (n= 1)
   Object = CASE 14

3 groups after combining group 0 (n = 8 ) and group 13 (n = 1) error =  13.897
Group 1 (n= 9)
   Object = CASE 1
   Object = CASE 2
   Object = CASE 3
   Object = CASE 6
   Object = CASE 7
   Object = CASE 14
```

```
   Object = CASE 15
   Object = CASE 16
   Object = CASE 17
Group 4 (n= 4)
   Object = CASE 4
   Object = CASE 8
   Object = CASE 9
   Object = CASE 20
Group 5 (n= 7)
   Object = CASE 5
   Object = CASE 10
   Object = CASE 11
   Object = CASE 12
   Object = CASE 13
   Object = CASE 18
   Object = CASE 19

2 groups after combining group 3 (n = 4 ) and group 4 (n = 7) error =  17.198
Group 1 (n= 9)
   Object = CASE 1
   Object = CASE 2
   Object = CASE 3
   Object = CASE 6
   Object = CASE 7
   Object = CASE 14
   Object = CASE 15
   Object = CASE 16
   Object = CASE 17
Group 4 (n= 11)
   Object = CASE 4
   Object = CASE 5
   Object = CASE 8
   Object = CASE 9
   Object = CASE 10
   Object = CASE 11
   Object = CASE 12
   Object = CASE 13
   Object = CASE 18
   Object = CASE 19
   Object = CASE 20

SCATTERPLOT - Plot of Error vs No. of Groups

                                                         Size of Error
                              |                             |- 18.06
                              |                             |- 17.20
                              |                             |- 16.34
   .                          |                             |- 15.48
                              |                             |- 14.62
                              |                             |- 13.76
                              |                             |- 12.90
      .                       |                             |- 12.04
                              |                             |- 11.18
                              |                             |- 10.32
------------------------------------------------------------|-  9.46
          .                   |                             |-  8.60
             .  .             |                             |-  7.74
                              |                             |-  6.88
                              |                             |-  6.02
                              |                             |-  5.16
                    .  .      |                             |-  4.30
                          .   |                             |-  3.44
                              *  .  .   .                   |-  2.58
                              |            .  .   .  .  .   |-  1.72
_______________________________________________________________
   |     |     |     |     |     |     |     |     |     |
No. of Groups
  2.00  4.00  6.00  8.00 10.00 12.00 14.00 16.00 18.00 20.00
```

Grouping Errors in Hierarchical Clustering

If you compare the results above with a discriminant analysis on the same data, you will see that the clustering procedure does not necessarily replicate the original groups. Clearly, “nearest neighbor” grouping in Euclidean space does not necessarily result in the same a priori groups from the discriminant analysis.

By examining the increase in error (variance of subjects within the groups) as a function of the number of groups, one can often make some decision about the number of groups one wishes to interpret. There is a large increase in error when going from 8 groups down to 7 in this analysis which suggests there are possibly 7 or 8 groups which might be examined. If we had more information on the objects of those groups, we might see a pattern or commonality shared by objects of those groups.