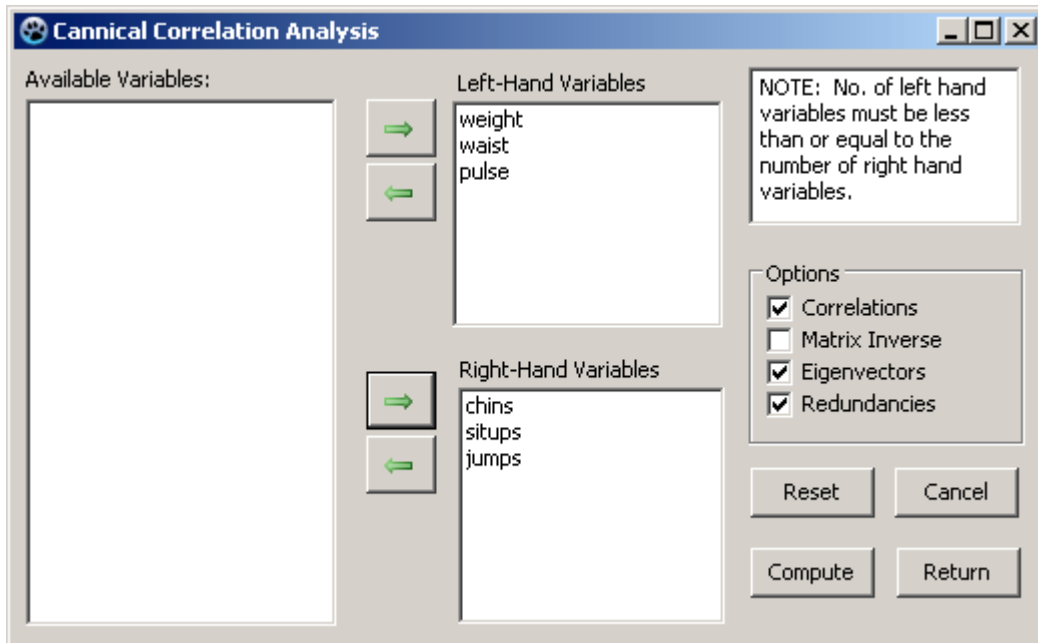# Canonical Correlation

An Example.

We will use the cancor.LAZ file to demonstrate canonical correlation analysis. This file consists of three measures of individuals (weight, waist size, pulse rate and three measures of physical ability, i.e. number of chin-ups, number of sit-ups and length of jumps. We are interested in obtaining the linear relationship between the sum of weighted subject characteristics (weight, waist and pulse) with the sum of weighted ability measures (chin-ups, sit-ups, jumps.) The dialog for our analysis is:

The results are obtained by clicking the Compute button. The results are:

```
CANONICAL CORRELATION ANALYSIS

Left Correlation Matrix with   20 cases.
Variables
                weight         waist          pulse
    weight       1.000         0.870         -0.366
     waist       0.870         1.000         -0.353
     pulse      -0.366        -0.353          1.000

Right Correlation Matrix with   20 cases.
Variables
                chins         situps          jumps
     chins       1.000         0.696          0.496
    situps       0.696         1.000          0.669
     jumps       0.496         0.669          1.000

Left-Right Correlation Matrix with   20 cases.
Variables
                chins         situps          jumps
    weight      -0.390        -0.493         -0.226
```

```
        waist       -0.552        -0.646        -0.191
        pulse        0.151         0.225         0.035

Right Inverse x Right-Left Matrix with   20 cases.
Variables
                    weight          waist          pulse
        chins       -0.102        -0.226          0.001
        situps      -0.552        -0.788          0.365
        jumps        0.193         0.448         -0.210

Left Inverse x Left-Right Matrix with   20 cases.
Variables
                     chins          situps         jumps
        weight       0.368          0.287         -0.259
        waist       -0.882         -0.890          0.015
        pulse       -0.026          0.016         -0.055

Canonical Function with   20 cases.
Variables
                    Var. 1         Var. 2         Var. 3
        Var. 1       0.162          0.172          0.023
        Var. 2       0.482          0.549          0.111
        Var. 3      -0.318         -0.346         -0.032

Trace of the matrix:=     0.6785
Percent of trace extracted:   100.0000

   Canonical R    Root   % Trace    Chi-Sqr     D.F.      Prob.
 1   0.795608     0.633  93.295     16.255        9      0.062
 2   0.200556     0.040   5.928      0.718        4      0.949
 3   0.072570     0.005   0.776      0.082        1      0.775

Overall Tests of Significance:
        Statistic        Approx. Stat.    Value    D.F.  Prob.>Value
Wilk's Lambda            Chi-Squared     17.3037     9    0.0442
Hotelling-Lawley Trace   F-Test          2.4938   9 38    0.0238
Pillai Trace             F-Test          1.5587   9 48    0.1551
Roys Largest Root        F-Test         10.9233   3 19    0.0002

Eigenvectors with   20 cases.
Variables
                    Var. 1         Var. 2         Var. 3
        Var. 1       0.210         -0.066          0.051
        Var. 2       0.635          0.022         -0.049
        Var. 3      -0.431          0.188          0.017

Standardized Right Side Weights with   20 cases.
Variables
                    Var. 1         Var. 2         Var. 3
        weight       0.775         -1.884          0.191
        waist       -1.579          1.181         -0.506
        pulse        0.059         -0.231         -1.051

Standardized Left Side Weights with   20 cases.
Variables
                    Var. 1         Var. 2         Var. 3
        chins        0.349         -0.376          1.297
```

```
      situps        1.054          0.123         -1.237
       jumps       -0.716          1.062          0.419

Raw Right Side Weights with   20 cases.
Variables
                    Var. 1         Var. 2         Var. 3
      weight        0.031         -0.076          0.008
       waist       -0.493          0.369         -0.158
       pulse        0.008         -0.032         -0.146

Raw Left Side Weights with   20 cases.
Variables
                    Var. 1         Var. 2         Var. 3
       chins        0.066         -0.071          0.245
      situps        0.017          0.002         -0.020
       jumps       -0.014          0.021          0.008

Right Side Correlations with Function with   20 cases.
Variables
                    Var. 1         Var. 2         Var. 3
      weight       -0.621         -0.772          0.135
       waist       -0.925         -0.378          0.031
       pulse        0.333          0.041         -0.942

Left Side Correlations with Function with   20 cases.
Variables
                    Var. 1         Var. 2         Var. 3
       chins        0.728          0.237          0.644
      situps        0.818          0.573         -0.054
       jumps        0.162          0.959          0.234

Redundancy Analysis for Right Side Variables
          Variance Prop.     Redundancy
       1    0.45080           0.28535
       2    0.24698           0.00993
       3    0.30222           0.00159

Redundancy Analysis for Left Side Variables
          Variance Prop.     Redundancy
       1    0.40814           0.25835
       2    0.43449           0.01748
       3    0.15737           0.00083
```

Interpreting The Standardized Canonical Coefficients.

The standardized weights are obtained from the characteristic equation. These elements are the coefficients with which to weight each of the standard (z) scores in our equation.

Typically, these weights are presented in two parts:
   a. The coefficients corresponding to each root are presented as column vectors for the left-hand weights and
   b. the coefficients corresponding to each root are presented as column vectors for the right-hand weights.

Structure Coefficents.

In addition to the standardized canonical coefficients, it is useful to obtain what are called structure coefficients. Structure coefficients are the correlations of the left-hand variables with the left-hand composite score (function) and the correlations of the right-hand variables with the right-hand function. They are obtained respectively as

We note from the results above that situps correlates the highest with the left-hand function in the first equation and that waist correlates highest with the right-hand function.

Redundancy Analysis

The proportion of variance obtained from the left-hand battery of variables from the canonical factor j (sum of weighted scores) is obtained.

Similarly, the proportion of variance obtained by the right-hand variables is obtained.

The redundancy of the left-hand variables given the availability of the right-hand variables which is displayed by canonical correlation j is obtained and the redundancy of the right-hand variables given the availability of the left-hand variables displayed by the canonical correlation j is obtained.

The total redundancy of the left-hand variables with the right-hand variables is simply

$$_t R_1 = \sum_{j=1}^{k} R_j$$

where k is the number of positive roots (rank of the canonical product matrix).

Similarly, the total redundancy of the right-hand variables with the left-hand variables is

$$_t R_r = \sum_{j=1}^{k} R_j$$

It should be noted that the left and right redundancy coefficients need not be equal, that is, it is possible that the left (or right) variables account for more variance of the right variables (or left) and vice versa.