

## Life Satisfaction Survey (continued) (Factor.sta)

This example is based on a fictitious data file (Factor.sta) describing a study of life satisfaction. This data file is also analyzed in the Example section of Factor Analysis.

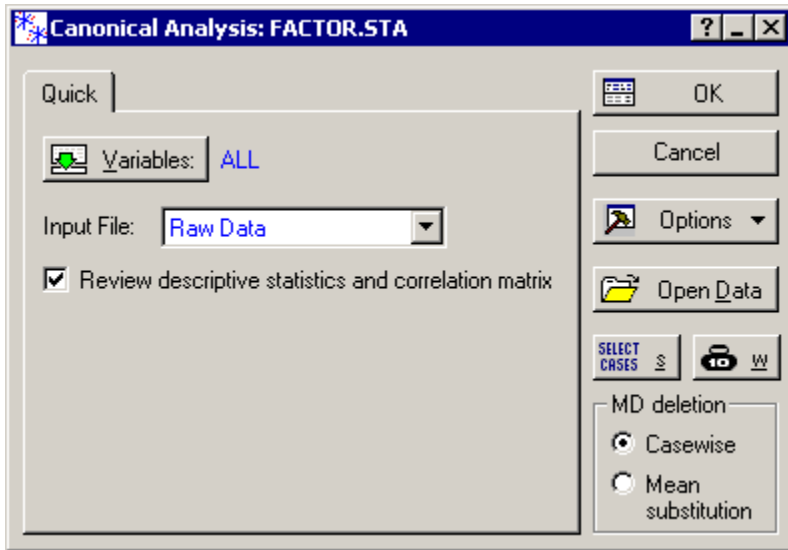
Suppose that a questionnaire is administered to a random sample of 100 adults. The questionnaire contains 10 items that are designed to measure satisfaction at work, satisfaction with hobbies (leisure time satisfaction), satisfaction at home, and general satisfaction in other areas of life. Responses to all questions are recorded via computer, and scaled so that the mean for all items is approximately 100.

The results are entered into the Factor.sta data file (see the partial listing shown below). Open this data file via the File - Open Examples menu; it is in the Datasets folder.

	1	2	3	4	5	6	7
	WORK 1	WORK 2	WORK 3	HOBBY 1	HOBBY 2	HOME 1	HOME 2
1	105.126	101.659	115.060	100.998	95.184	100.281	101.667
2	77.049	72.933	77.485	72.744	61.563	93.854	95.392
3	86.017	82.206	78.889	77.951	91.705	86.773	108.070
4	91.425	106.107	95.640	90.901	111.466	100.248	86.080
5	113.714	92.029	99.079	79.277	98.416	104.013	83.271
6	86.606	87.817	67.663	93.662	77.997	99.822	97.275
7	95.067	94.505	98.081	94.513	97.422	93.694	99.181
8	113.500	104.607	105.572	101.008	102.275	87.427	96.664
9	104.549	97.299	94.074	88.538	98.112	97.785	99.585
10	104.635	97.908	85.823	82.486	90.447	104.688	95.076
11	102.064	87.010	94.687	79.203	68.482	78.995	86.430
12	109.428	94.937	104.396	119.293	112.988	122.931	114.816
13	89.994	77.392	100.771	97.026	111.107	107.660	116.858
14	89.983	91.259	89.216	81.250	100.519	94.902	96.723
15	96.946	103.043	97.493	98.515	96.833	92.778	110.072
16	93.210	93.843	87.345	104.256	115.323	114.286	111.617
17	70.905	65.903	74.231	60.108	47.088	79.806	94.799
18	114.854	98.083	105.634	90.237	101.011	98.951	87.921
19	100.315	92.518	87.609	103.273	88.647	108.401	103.565
20	111.355	104.373	101.857	108.093	121.894	99.054	96.688

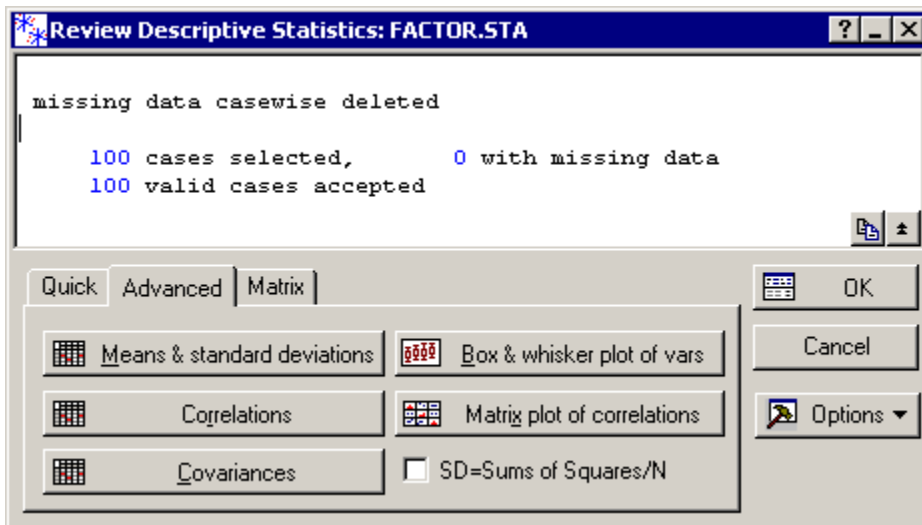
**Purpose of Analysis.** Suppose that you want to learn about the relationship of work satisfaction to satisfaction in other domains. Conceptually, you will treat the work satisfaction items as the explanatory or independent variables, and the other satisfaction items as the dependent variables.

**Initial Computations.** From the Statistics - Multivariate Exploratory Techniques menu, select Canonical Analysis to display the Canonical Analysis Startup Panel. Click the Variables button on the Startup Panel to display the Select variables for canonical analysis dialog. Here, select all variables in this file and click the OK button. The Startup Panel will now look like this:



Canonical correlation analysis is based on the correlation matrix of variables. Therefore, the first step of the analysis is to compute that correlation matrix (unless a Correlation Matrix input file is specified via the Input File drop-down box, in which case the input needs to be a correlation matrix). Note that you can later select variables for the analysis (for the two sets) from among those that are specified at this point.

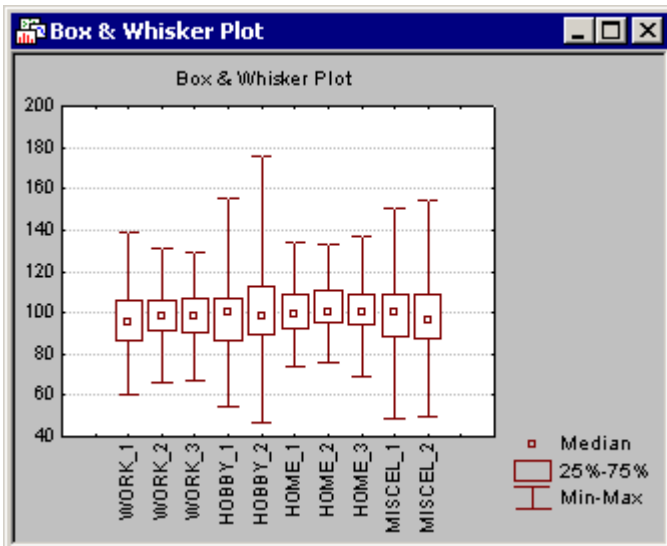
Select the Review descriptive statistics and correlation matrix check box in order to compute the detailed descriptive statistics (i.e., means, correlations, covariances) for the variables in the current analysis. Now, click the OK button to display the Review Descriptive Statistics dialog.



Plots. In order to visualize the distribution of the variables, two types of plots are available from the Review Descriptive Statistics - Advanced tab: Box & whisker plot of vars and Matrix plot of correlations.

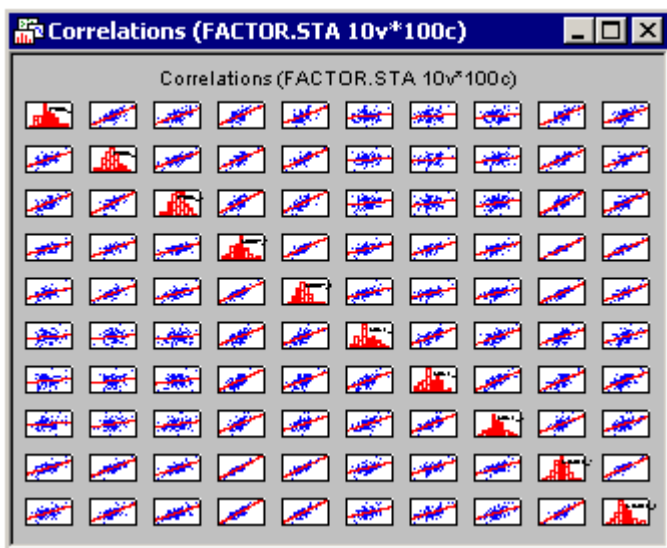
Click the Box & whisker plot of vars button, select all variables in the data set in the resulting variable selection dialog, and click the OK button. The Box-Whisker Type dialog will then be displayed in which you can select from four types of box and whisker plots (for this example, select the Median/Quart./Range option button and then click the OK button). The central tendency (e.g., median) and variability (e.g., quartile and range) of the selected variables are displayed in these plots (note that

the specific method for computing the quartiles and the median can be specified on the Options - General tab).



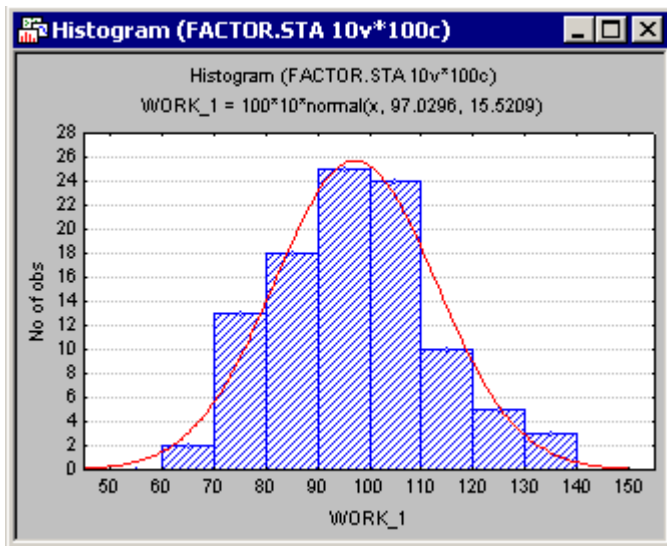
Box & whisker plots are useful to determine if the distribution of a variable is symmetrical. If the distribution is not symmetrical, then you may want to view the histogram for the respective variable.

Now, click the Matrix plot of correlations button (and select all of the variables in the variable selection dialog) to display a matrix of scatterplots. These plots should be examined for outliers, which may greatly bias the computation of the correlation coefficients, and thus the canonical analysis (see the discussion in the Introductory Overview).



Next, click the Means & standard deviations button on the Review Descriptive Statistics - Advanced tab to display a spreadsheet containing the means and standard deviations of the selected variables. As in most other modules, the default plot for the spreadsheet of means and standard deviations is the histogram of the distribution of the respective variable. This histogram will show the normal curve superimposed over the observed distribution, to provide a visual check for any violations of the normality

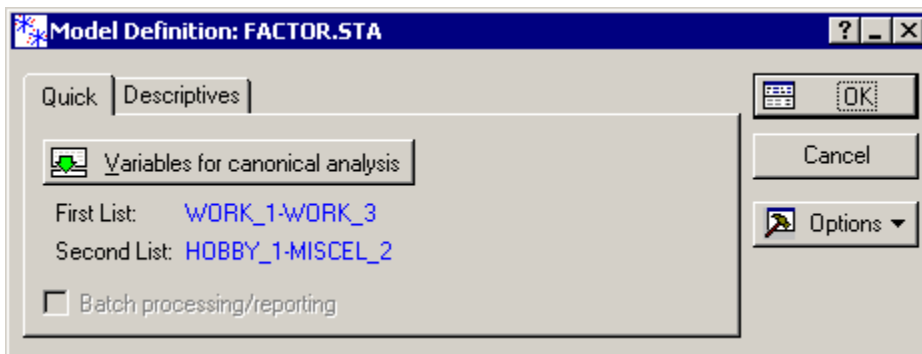
assumption. For example, to produce the histogram for variable Work\_1, right-click the Mean for variable Work\_1 (97.0296) and then select Graphs of Input Data - Histogram Work\_1 - Normal Fit in the resulting shortcut menu to produce the graph shown below.



The distribution of this variable (of subjects' responses to the first item) follows the normal distribution. Thus, there is little reason to suspect that this variable violates the normality assumption.

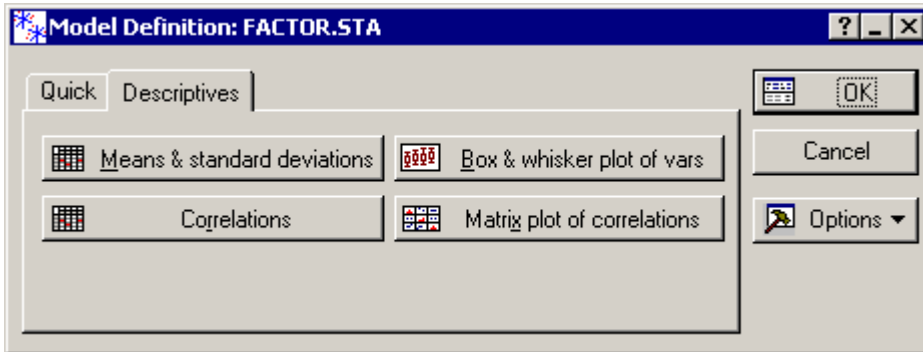
Specifying the Canonical Analysis. To proceed with the canonical correlation analysis, click the OK button to exit the Review Descriptive Statistics dialog and display the Model Definition dialog in which you can select the left and right set variables.

Specifying variables. To specify variables for the two sets, click the Variables for canonical analysis button on the Model Definition - Quick tab to open the standard variable selection window. Select the work satisfaction items (i.e., variables Work\_1, Work\_2, and Work\_3) in the First variable list, and the remaining satisfaction items (i.e., variables Hobby\_1, Hobby\_2, Home\_1, Home\_2, Home\_3, Miscel\_1, Miscel\_2) in the Second variable list.

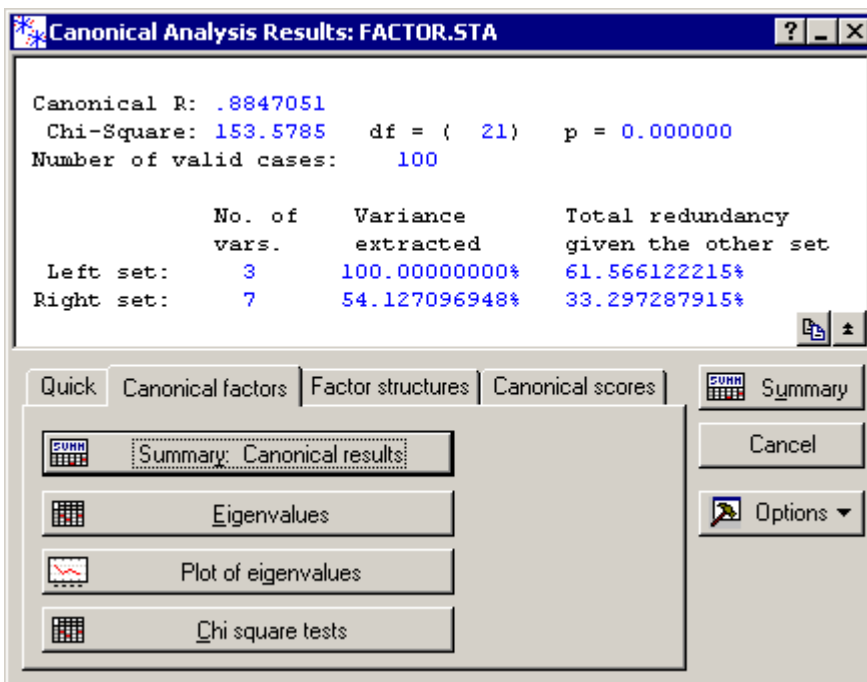


Note that the designation of first and second list is arbitrary here; that is, you could also select the work satisfaction items into the second list and the remaining satisfaction items into the first list. In a sense, canonical analysis is completely "symmetrical," that is, it will compute the same statistics (loadings, weights, etc.) for the variables in both lists.

You may also use some other options, such as the Means & standard deviations or Correlations buttons, as well as create some useful descriptive graphs via the Model Definition - Descriptives tab.



Reviewing Results. After specifying the two lists of variables, you are now ready to begin the analysis. Click the OK button on the Model Definition dialog. After a few moments, the Canonical Analysis Results dialog will be displayed.



The nature of most of the statistics reported here in the results summary are reviewed in the Introductory Overview section of this module. Therefore, the focus in this example will be on the interpretation of the results.

Click the Summary: Canonical results button on the Quick tab or the Canonical factors tab to display the Canonical Analysis Summary spreadsheet.

Data: Canonical Analysis Summary (FACTOR.STA)*		
Canonical Analysis Summary (FACTOR.STA) Canonical R: .88471 Chi <sup>2</sup> (21)=153.58 p=0.0000		
N=100	<b>Left Set</b>	<b>Right Set</b>
<b>No. of variables</b>	3	7
Variance extracted	100.000%	54.1271%
Total redundancy	61.5661%	33.2973%
<b>Variables:</b>		
1	WORK_1	HOBBY_1
2	WORK_2	HOBBY_2
3	WORK_3	HOME_1
4		HOME_2
5		HOME_3
6		MISCEL_1
7		MISCEL_2

**Canonical R.** The overall canonical R is fairly substantial (.88), and highly significant ( $p < .001$ ). Remember that the canonical R reported here pertains to the first and most significant canonical root. Thus, this value can be interpreted as the simple correlation between the weighted sum scores in each set, with the weights pertaining to the first (and most significant) canonical root.

**Overall Redundancy.** The values in the rows labeled Variance extracted and Total redundancy give an indication of the magnitude of the overall correlations between the two sets of variables, relative to the variance of the variables. This is different from the canonical R-square, because the latter statistic expresses the proportion of variance accounted for in the canonical variates (see the Introductory Overview for further details).

**Variance extracted.** The values in the Variance extracted row indicate the average amount of variance extracted from the variables in the respective set by all canonical roots. Thus, all three roots extract 100% of the variance from the left set, that is, the three work satisfaction items, and 54% of the variance in the right set. Note that one of these values will always be 100% because STATISTICA extracts as many roots as the minimum number of variables in either set. Thus, for one set of variables, there are as many independent sums (canonical variates) as there are variables. Intuitively, it should be clear that, for example, three independent sum scores derived from three variables will explain 100% of all variability.

**Redundancy.** The computation of the Total redundancy is explained in the Introductory Overview section. These values can be interpreted such that, based on all canonical roots, given the right set of variables (the seven non-work related satisfaction items), you can account for, on the average, 61.6% of the variance in the variables in the left set (the satisfaction items). Likewise, you can account for 33.3% of the variance in the non-work related satisfaction items, given the work related satisfaction items. These results suggest a fairly strong overall relationship between the items in the two sets.

**Testing the Significance of Canonical Roots.** Now, check whether all three canonical roots are significant. Keep in mind that the canonical R reported in this spreadsheet represents only the first root, that is, the strongest and most significant canonical correlation. To test the significance of all canonical roots, click the Chi square tests button on the Canonical factors tab of the Canonical Analysis Results dialog. These results will be displayed in Chi-Square with Successive Roots Removed spreadsheet.

Data: Chi-Square Tests with Successive Roots Removed (FACTOR.STA)*						
Root Removed	Chi-Square Tests with Successive Roots Removed (FACTOR.STA)					
	Canonical R	Canonical R-sqr.	Chi-sqr.	df	p	Lambda Prime
0	0.884705	0.782703	153.5785	21	0.000000	0.193486
1	0.271080	0.073484	10.8516	12	0.541689	0.890422
2	0.197374	0.038957	3.7153	5	0.591097	0.961043

The maximum number of roots that can be extracted is equal to the smallest number of variables in either set. Since three work satisfaction items were selected into the first set, STATISTICA would be expected to extract three canonical roots.

The sequential significance test works as follows. First, look at all three canonical variables together, that is, without any roots removed. That test is highly significant. Next, the first (and, as you know, most significant) root is "removed" and the statistical significance of the remaining two roots is determined. That test (in the second row of the spreadsheet) is not significant. You can stop at this point and conclude that only the first canonical root is statistically significant, and it should be examine further. If the second test were also statistically significant, you would then proceed to the third line of the spreadsheet to see whether the remaining third canonical root is also significant.

Factor Structure and Redundancy. You now know that you should consider further only the first canonical root. How can this root be interpreted, that is, how is it correlated with the variables in the two sets? As discussed in the Introductory Overview, the interpretation of canonical "factors" follows a similar logic to that in Factor Analysis. Specifically, you can compute the correlations between the items in each set with the respective canonical root or variable (remember that the canonical variable in each set is "created" as the weighted sum of the variables). Those correlations are also called canonical factor loadings or structure coefficients.

You can compute those values (as well as the variance extracted for each set) via the Factor structures & redundancies button of the Factor structures tab on the Canonical Analysis Results dialog. Click this button to generate four results spreadsheets; namely (1) Factor Structure, left set, (2) Variance Extracted (Proportions), left set, (3) Factor Structure, right set, and (4) Variance Extracted (Proportions), right set.

Factor structure in the left set. First, examine the loadings for the left set.

Data: Factor Structure, left set (FACTOR.STA)*			
Variable	Factor Structure, left set (FACTOR.STA)		
	Root 1	Root 2	Root 3
WORK_1	0.796461	-0.575348	0.186075
WORK_2	0.952643	0.100725	-0.286926
WORK_3	0.875390	0.217192	0.431880

Remember that only the first canonical root is statistically significant, and it is the only one that should be interpreted. As you can see, the three work satisfaction items show substantial loadings on the first canonical factor, that is, they correlate highly with that factor.

As a measure of redundancy, the average amount of variance accounted for in each item by the first root could be computed. To do so, you could sum up the squared canonical factor loadings and divide them by 3 (the number of variables in this set). Click the Factor structures & redundancies button on the Canonical Analysis Results - Factor structures tab to obtain various spreadsheets. First, look at the Variance Extracted (Proportions), left set results spreadsheet.

Data: Variance Extracted (Proportions), left set (FACTOR.STA)*		
Factor	Variance Extracted (Proportions), left set (FACTOR.STA)	
	Variance extracted	Reddncy.
Root 1	0.769395	0.602208
Root 2	0.129448	0.009512
Root 3	0.101157	0.003941

As you can see in this spreadsheet, the first canonical root extracts an average of about 77% of the variance from the work satisfaction items. If you multiply that value with the proportion of shared variance between the canonical variates in the two sets (i.e., with R-square), then the number in the Reddncy column of the spreadsheet (i.e., redundancy) is obtained. Thus, given the variables in the right set (the non-work related satisfaction items), you can account for about 60% of the variance in the work related satisfaction items, based on the first canonical root.

Factor structure in right set. In the Factor Structure, right set spreadsheet, the first canonical root or factor is marked by high loadings on the leisure satisfaction items (Hobby\_1 and Hobby\_2).

Data: Factor Structure, right set (FACTOR.STA)*			
Variable	Factor Structure, right set (FACTOR.STA)		
	Root 1	Root 2	Root 3
HOBBY_1	0.823914	-0.099577	0.055376
HOBBY_2	0.808859	0.277944	-0.149747
HOME_1	0.186554	-0.017177	0.236525
HOME_2	0.238228	0.253169	0.456321
HOME_3	0.278220	0.414321	0.173396
MISCEL_1	0.861283	0.048749	0.328389
MISCEL_2	0.820765	0.245223	0.241010

The loadings are much lower for the home-related satisfaction items. Therefore, you can conclude that the significant canonical correlation between the variables in the two sets (based on the first root) is probably the result of a relationship between work satisfaction, and leisure time and general satisfaction. If you consider work satisfaction as the explanatory variable, you could say that work satisfaction affects leisure time and general satisfaction, but not (or much less so) satisfaction with home life.

The Variance Extracted (Proportions), right set spreadsheet shows the redundancies for the right set of variables.

Data: Variance Extracted (Proportions), right set (FACTOR.STA)...		
Variable	Variance Extracted (Proportions), right set (FACTOR.STA)	
	Variance extractd	Reddncy.
Root 1	0.416787	0.326221
Root 2	0.055104	0.004049
Root 3	0.069379	0.002703

As you can see, the first canonical root accounts for an average of roughly 42% of variance in the variables in the right set; given the work satisfaction items, you can account for about 33% of the variance in the other satisfaction items, based on the first canonical root. Note that these numbers are "pulled down" by the relative lack of correlations between this canonical variate and the home satisfaction items.

Canonical Scores. Remember that the canonical variates represent weighted sums of the variables in each set. You can review those weights by using the Canonical Analysis Results - Canonical Scores tab on the dialog.

Click the Left & right set canonical weights button to display the following separate spreadsheets:

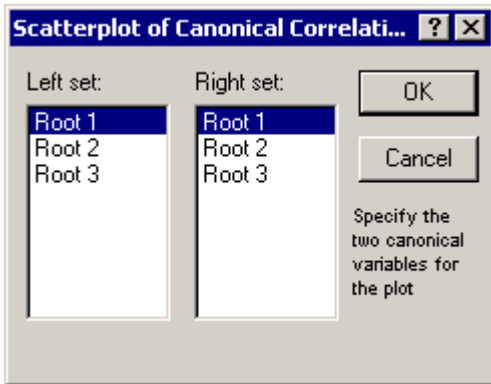
Data: Canonical Weights, left set (FACTOR.STA)*			
Variable	Canonical Weights, left set (FACTOR.STA)		
	Root 1	Root 2	Root 3
WORK_1	0.217021	-1.35890	0.24350
WORK_2	0.592485	0.37139	-1.38769
WORK_3	0.300124	0.83222	1.28861

Data: Canonical Weights, right set (FACTOR.STA)			
Variable	Canonical Weights, right set (FACTOR.STA)		
	Root 1	Root 2	Root 3
HOBBY_1	0.249966	-1.63580	-1.29069
HOBBY_2	0.221850	1.00467	-1.09591
HOME_1	-0.156475	-0.43401	0.30966
HOME_2	-0.447900	0.32580	0.45678
HOME_3	-0.066126	0.69965	-0.31653
MISCEL_1	0.615556	-0.36114	1.73651
MISCEL_2	0.290859	0.79787	0.45774

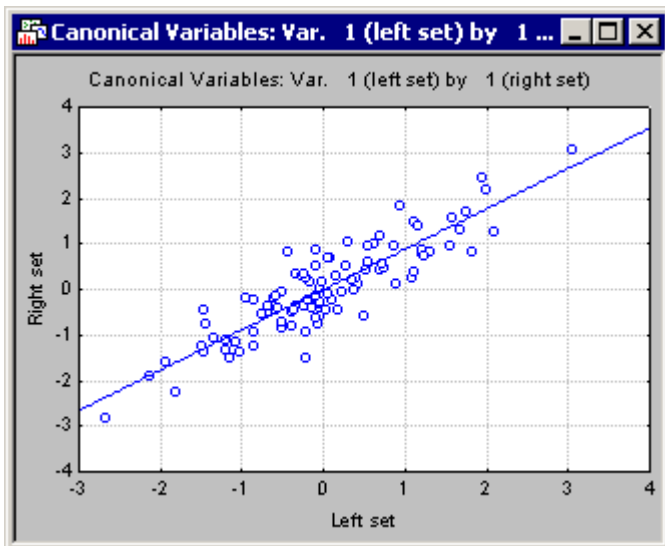
The weights shown in the Canonical Weights spreadsheets pertain to the standardized (z-transformed) variables in the two sets. You can use those weights to compute scores for the canonical variates. The scores computed from the data in the current data file may be saved via the Save canonical scores button on the Canonical Analysis Results - Canonical Scores tab.

Plotting canonical scores. Now, plot the canonical scores for the variables in the left set against the scores for the variables in the right set. Click the Scatterplot of canonical correlations button on the Canonical

Analysis Results - Factor structures tab to display the Scatterplot of Canonical Correlations dialog. To produce a scatterplot for the first (and only significant) canonical variate, select Root 1 in the Left set box and Root 1 in the Right set box.



Now, click the OK button to produce the desired scatterplot. Note that in the illustration shown below a linear regression line was added to the plot via the Plot: Fitting options pane of Graph Options dialog.



There are no outliers apparent in this plot, nor do the residuals around the regression line indicate any non-linear trend (e.g., by forming a U or S around the regression line). Therefore, you can be satisfied that no major violations of a main assumption of canonical correlation analysis are evident.

Clusters of cases. Another interesting aspect of this plot is whether or not there is any evidence of clustering of cases. Such clustering may happen if the sample is somehow heterogeneous in nature. For example, suppose respondents from two very different industries who are working under very different conditions were included in the sample. It is conceivable that the canonical correlation represented by the plot above could then be the result of the fact that one group of respondents is generally more satisfied with their work and leisure time (and life in general). If so, this would be reflected in this plot by two distinct clusters of points: one at the low ends of both axes and one at the high ends. However, in this example, there is no evidence of any natural grouping of this kind, and therefore you do not have to be concerned.

Conclusion. It can be concluded from the analysis of this (fictitious) dataset that satisfaction at work affects leisure time satisfaction and general satisfaction. Satisfaction with home life did not seem to be affected. In practice, before generalizing these conclusions, you should replicate the study. Specifically, you should ensure that the canonical factor structure that led to the interpretation of the first canonical root is reliable (i.e., replicable).