

The very first thing to determine is if there is any relationship between the two sets of variables at all. This is carried out using Wilk's lambda. The results of this are found on page 1 of the output of the SAS Program.

|  | **Canonical Correlation** | **Adjusted Canonical Correlation** | **Approximate Standard Error** | **Squared Canonical Correlation** | **Test of H0: The canonical correlations in the current row and all that follow are zero** | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Likelihood Ratio** | **Approximate F Value** | **Num DF** | **Den DF** | **Pr > F** |
| **1** | 0.917329 | 0.908257 | 0.020295 | 0.841493 | 0.04859891 | 17.50 | 16 | 165.61 | <.0001 |
| **2** | 0.816927 | 0.809569 | 0.042589 | 0.667370 | 0.30660427 | 9.31 | 9 | 134.01 | <.0001 |
| **3** | 0.265385 | 0.205837 | 0.119019 | 0.070429 | 0.92175668 | 1.16 | 4 | 112 | 0.3305 |
| **4** | 0.091684 | . | 0.126961 | 0.008406 | 0.99159404 | 0.48 | 1 | 57 | 0.4898 |

From this table:

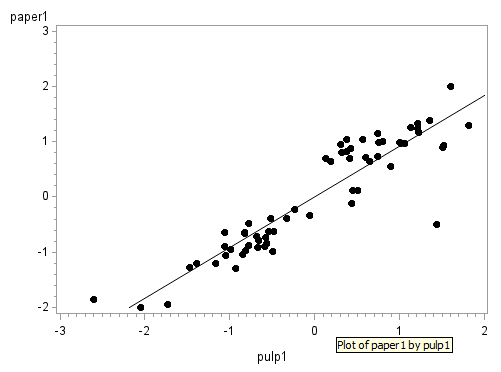
|  |  |  |
| --- | --- | --- |
| Hypothesis (canonical correlations for ρ) | P value | Conclusion |
| H0:ρ1\*=ρ2\*=⋯=ρp\*=0 | <.0001 | Reject NULL and ρ1\* ≠ 0 |
| H0:ρ2\*=ρ3\*= ρ4\*=0 | <.0001 | Reject NULL and ρ2\* ≠ 0 |
| H0:ρ3\*= ρ4\*=0 | 0.3305 | Fail to reject null (STOP here) |
| Further evaluation not required |  |  |

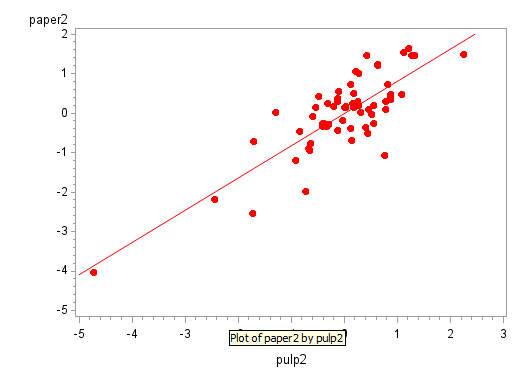
Therefore first two canonical variate pairs are significantly correlated and dependent on one another.

Now that we have tested the hypotheses of independence and have rejected them, the next step is to obtain estimates of canonical correlation.

From the above table we have the following interpretation:

* We see that 84.15% of the variation in U1 is explained by the variation in V1, and 66.74% of the variation in U2 is explained by V2
* We don’t go any further since we were not able to reject the null





SAS output provides the estimated canonical coefficients (aij) for the paper variables which are provided in the following table.

| **Raw Canonical Coefficients for the Paper Variables** | | | | |
| --- | --- | --- | --- | --- |
|  | **paper1** | **paper2** | **paper3** | **paper4** |
| **breaking\_length** | -0.52244265 | -1.213137146 | -1.978680694 | -1.764664583 |
| **elastic\_modulus** | -0.29578986 | -2.153646217 | 4.9200568562 | -0.818894587 |
| **stress\_at\_failure** | 1.3660327595 | 0.7355096103 | -3.222026651 | 4.1489040115 |
| **burst\_strength** | 0.9760405051 | 5.4369965657 | 10.321875087 | -0.98997032 |

Thus, using the coefficient values in the first column, the first canonical variable for paper can be determined using the following formula:

U1= -0.522 \* Xbreaking\_length - 0.296 \* Xelastic\_modulus + 1.366 \* Xstress\_at\_failure + 0.976 \* Xburst\_strength

Likewise, the estimated canonical coefficients (bij) for the pulp are located in the next table in the SAS output:

| **Raw Canonical Coefficients for the Pulp Variables** | | | | |
| --- | --- | --- | --- | --- |
|  | **pulp1** | **pulp2** | **pulp3** | **pulp4** |
| **arithmetic\_fiber\_length** | -0.63837718 | 2.759452713 | -2.055673602 | 9.3494101482 |
| **long\_fiber\_fraction** | 0.0425404867 | 0.0674575788 | 0.0051919967 | -0.146642731 |
| **fine\_fiber\_fraction** | 0.0185013292 | 0.0002853256 | -0.094704066 | 0.0012658215 |
| **zero\_span\_tensile** | 27.733124475 | -52.95920077 | -26.40008876 | 3.0226089805 |

U1= -0.638 \* Xarithmetic\_fiber\_length + 0.0425 \* Xlong\_fiber\_fraction + 0.185 \* Xfine\_fiber\_fraction + 27.733 \* Xzero\_span\_tensile

**The correlations between each variable and the corresponding canonical variate**

1. The correlations between the paper variables and the canonical variables for Paper are:

| **Correlations Between the Paper Variables and Their Canonical Variables** | | | | |
| --- | --- | --- | --- | --- |
|  | **paper1** | **paper2** | **paper3** | **paper4** |
| **breaking\_length** | 0.9351 | -0.1261 | -0.0534 | -0.3270 |
| **elastic\_modulus** | 0.8869 | -0.4280 | 0.1306 | -0.1148 |
| **stress\_at\_failure** | 0.9767 | -0.1453 | -0.0307 | -0.1549 |
| **burst\_strength** | 0.9518 | 0.0147 | 0.0127 | -0.3061 |

Looking at the first canonical variable for paper, we see that all correlations are uniformly large. Therefore, you can think of this canonical variate as an overall measure of Paper Characteristics. For the second canonical variable for Paper, only elastic\_modulus is moderately large.

1. The correlations between the pulp variables and the canonical variables for pulp are:

| **Correlations Between the Pulp Variables and Their Canonical Variables** | | | | |
| --- | --- | --- | --- | --- |
|  | **pulp1** | **pulp2** | **pulp3** | **pulp4** |
| **arithmetic\_fiber\_length** | 0.8166 | 0.3683 | 0.1661 | 0.4122 |
| **long\_fiber\_fraction** | 0.9056 | 0.3848 | 0.1779 | -0.0126 |
| **fine\_fiber\_fraction** | -0.6496 | 0.0123 | -0.7309 | -0.2087 |
| **zero\_span\_tensile** | 0.9395 | -0.2307 | 0.1851 | 0.1730 |

Looking at the first canonical variable for pulp, we see that all correlations are uniformly large. Therefore, you can think of this canonical variate as an overall measure of Pulp Fiber Characteristics. Most of the correlations with the second canonical variable are small.

To reinforce the above results we look at the correlations between each set of variables and the opposite group of canonical variates

| **Correlations Between the Paper Variables and the Canonical Variables of the Pulp Variables** | | | | |
| --- | --- | --- | --- | --- |
|  | **pulp1** | **pulp2** | **pulp3** | **pulp4** |
| **breaking\_length** | 0.8578 | -0.1030 | -0.0142 | -0.0300 |
| **elastic\_modulus** | 0.8136 | -0.3496 | 0.0346 | -0.0105 |
| **stress\_at\_failure** | 0.8960 | -0.1187 | -0.0081 | -0.0142 |
| **burst\_strength** | 0.8731 | 0.0120 | 0.0034 | -0.0281 |

We can see that all three of these correlations are strong and show a pattern similar to that with the canonical variate for Paper. The reason for this is obvious: The first canonical correlation is very high.

| **Correlations Between the Pulp Variables and the Canonical Variables of the Paper Variables** | | | | |
| --- | --- | --- | --- | --- |
|  | **paper1** | **paper2** | **paper3** | **paper4** |
| **arithmetic\_fiber\_length** | 0.7491 | 0.3009 | 0.0441 | 0.0378 |
| **long\_fiber\_fraction** | 0.8307 | 0.3144 | 0.0472 | -0.0012 |
| **fine\_fiber\_fraction** | -0.5959 | 0.0100 | -0.1940 | -0.0191 |
| **zero\_span\_tensile** | 0.8618 | -0.1885 | 0.0491 | 0.0159 |

Note that these also show a pattern similar to that with the canonical variate for pulp. Again, this is because the first canonical correlation is very high.



Data: sales.txt with: Sales Growth, Sales Profitability, New Account Sales, Creativity, Mechanical Reasoning, Abstract Reasoning, Mathematics

We have the following table from SAS:

Canonical Correlation Analysis

|  | **Canonical Correlation** | **Adjusted Canonical Correlation** | **Approximate Standard Error** | **Squared Canonical Correlation** | **Test of H0: The canonical correlations in the current row and all that follow are zero** | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Likelihood Ratio** | **Approximate F Value** | **Num DF** | **Den DF** | **Pr > F** |
| **1** | 0.984736 | 0.984079 | 0.004328 | 0.969706 | 0.01523815 | 106.51 | 6 | 90 | <.0001 |
| **2** | 0.704977 | 0.701779 | 0.071858 | 0.496993 | 0.50300709 | 22.73 | 2 | 46 | <.0001 |

1. The value of the first canonical correlation is 0.985 and its square is 0.97. The squared value is interpreted as: 97% of the variation in U1 is explained by the variation in V1
2. We see the following tables:

| **Raw Canonical Coefficients for the Sales Variables** | | |
| --- | --- | --- |
|  | **sales1** | **sales2** |
| **growth** | 0.0365113335 | -0.359335905 |
| **profit** | 0.0737615658 | 0.2511530404 |

Thus, using the coefficient values in the first column, the first canonical variable for sales can be determined using the following formula:

U1 = 0.037 \* Xgrowth + 0.074 \* Xprofit

| **Raw Canonical Coefficients for the Test Scores** | | |
| --- | --- | --- |
|  | **scores1** | **scores2** |
| **mech** | 0.0902480689 | 0.0666925352 |
| **abs** | -0.021027596 | -0.565702124 |
| **math** | 0.0776605103 | 0.0468490918 |

Thus, using the coefficient values in the first column, the first canonical variable for test scores can be determined using a similar formula:

V1 = 0.090 \* Ymech - 0.021 \* Yabstract + 0.078 \* Ymath

1. With the following values: sales growth = 93, profitability = 96, mechanical reasoning = 12, abstract reasoning = 10 and math ability = 20

U1 = 0.037 \* Xgrowth + 0.074 \* Xprofit = 0.037 \* 93 + 0.074 \* 96 = 10.545

V1 = 0.090 \* Ymech - 0.021 \* Yabstract + 0.078 \* Ymath = 0.090 \* 12 - 0.021 \* 10 + 0.078 \* 20 = 2.43

The first pair of unstandardized canonical variables is (10.545, 2.43)

1. We have the following table extracted from the SAS output:

|  |  |  |
| --- | --- | --- |
| Hypothesis (canonical correlations for ρ) | Likelihood ration, F value, Num DF and Den DF, P value | Conclusion |
| H0:ρ1\*=ρ2\*=0 | 0.0152, 106.51, 6 and 90, <.0001 | Reject NULL and ρ1\* ≠ 0 |
| H0:ρ2\*= 0 | 0.503, 22.73, 2 and 46, <.0001 | Reject NULL and ρ2\* ≠ 0 |

From the above table we see Wilks’ lambda Λ = 0.0152; F = 106.51; d.f. = 6, 90; p < 0.0001. Wilks' lambda is ratio of two variance-covariance matrices (raised to a certain power). If the values of these statistics are too large (small p-value), it indicates rejection of the null hypothesis.

Here we reject the null hypothesis that there is no relationship between the two sets of variables, and can conclude that the two sets of variables are dependent.

Since Wilk's lambda is significant, and since the canonical correlations are ordered from largest to smallest, we can conclude that at least ρ1\*≠0.

1. From the above table we see Wilks’ lambda Λ = 0.503; F = 22.73; d.f. = 2, 46; p < 0.0001.

This is also significant so we reject the null and conclude that the second canonical variate pair is correlated.

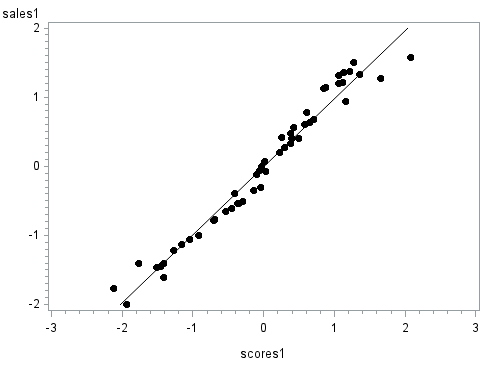
1. The highlighted section of the table below gives the correlations between the sales variables (the x-variables) and the first canonical variable that is constructed from them

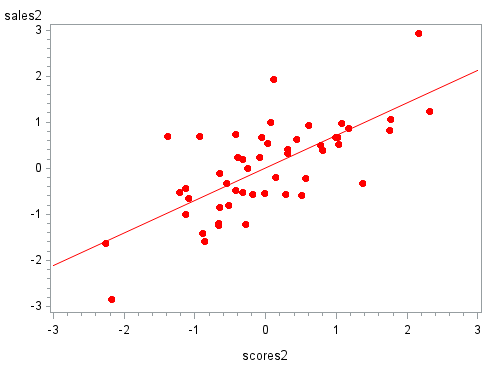
| **Correlations Between the Sales Variables and Their Canonical Variables** | | |
| --- | --- | --- |
|  | **sales1** | **sales2** |
| **growth** | 0.9595 | -0.2818 |
| **profit** | 0.9949 | 0.1011 |

1. The highlighted section of the table below gives the correlations between the test variables (the y-variables) and the first canonical variable that is constructed from them

| **Correlations Between the Test Scores and Their Canonical Variables** | | |
| --- | --- | --- |
|  | **scores1** | **scores2** |
| **mech** | 0.7583 | 0.0422 |
| **abs** | 0.5364 | -0.8437 |
| **math** | 0.9684 | -0.0621 |

1. We have two canonical correlations only because there are only p = 2 variables in the first group relating to Sale Performance.
2. To interpret the relationship we know from above that both the canonical variate pairs are correlated. Further we see that 97% of the variation in U1 is explained by the variation in V1, and 49.7% of the variation in U2 is explained by V2. While the first one is very high, the second is also moderately high and implies that only the both the canonical correlations are important.





To interpret each component, we must compute the correlations between each variable and the corresponding canonical variate. The correlations between the sales variables and the canonical variables for Sales Performance are:

| **Correlations Between the Sales Variables and Their Canonical Variables** | | |
| --- | --- | --- |
|  | **sales1** | **sales2** |
| **growth** | 0.9595 | -0.2818 |
| **profit** | 0.9949 | 0.1011 |

Looking at the first canonical variable for sales, we see that all correlations are uniformly large. Therefore, you can think of this canonical variate as an overall measure of Sales Performance. For the second canonical variable for Sales Performance, none of the correlations is particularly large, and so, this canonical variable yields little information about the data.

The correlations between the test scores and the canonical variables for Test Scores are

| **Correlations Between the Test Scores and Their Canonical Variables** | | |
| --- | --- | --- |
|  | **scores1** | **scores2** |
| **mech** | 0.7583 | 0.0422 |
| **abs** | 0.5364 | -0.8437 |
| **math** | 0.9684 | -0.0621 |

Since all correlations are large for the first canonical variable, this can be thought of as an overall measure of test performance as well, however, it is most strongly correlated with mathematics test scores.

The correlations with the second canonical variable are small except with abstract reasoning where we have negative correlation.

Putting the two outcomes together, we see that the best predictor of sales performance is mathematics test scores as this indicator stands out most.

In order to reinforce these results we look at the correlations between each set of variables and the opposite group of canonical variates.

The correlations between the sales variables and the canonical variates for test scores are:

| **Correlations Between the Sales Variables and the Canonical Variables of the Test Scores** | | |
| --- | --- | --- |
|  | **scores1** | **scores2** |
| **growth** | 0.9448 | -0.1987 |
| **Profit** | 0.9797 | 0.0713 |

We can see that both of these correlations are strong and show a pattern similar to that with the canonical variate for sales. The reason for this is obvious: The first canonical correlation is very high.

The correlations between the test and the canonical variates for sales are:

| **Correlations Between the Test Scores and the Canonical Variables of the Sales Variables** | | |
| --- | --- | --- |
|  | **sales1** | **sales2** |
| **mech** | 0.7467 | 0.0298 |
| **abs** | 0.5282 | -0.5948 |
| **math** | 0.9536 | -0.0438 |

Note that these also show a pattern similar to that with the canonical variate for test scores. Again, this is because the first canonical correlation is very high.

These results confirm that sales performance is best predicted by mathematics test scores.