What is the procedure for conducting Canonical Correlation Analysis in SPSS?

Authored by

stats writer

June 29, 2024

RECOMMENDED CITATION

stats writer (2024). *What is the procedure for conducting Canonical Correlation Analysis in SPSS?*. PSYCHOLOGICAL SCALES. Retrieved from https://scales.arabpsychology.com/?p=158812

Canonical Correlation Analysis (CCA) is a statistical technique used for exploring the relationship between two sets of variables. It is commonly used to identify the underlying patterns and associations between two different data sets. In SPSS, the procedure for conducting CCA involves the following steps:

1. Data preparation: Before performing CCA, the two sets of variables must be organized into separate data files.

2. Accessing the CCA tool: In SPSS, CCA can be found under the "Analyze" menu, under the "Multivariate" option.

3. Selecting variables: In the CCA tool, the user must specify which variables from each data set will be used in the analysis.

4. Setting options: SPSS offers various options for CCA, such as the type of correlation matrix to be used and the method for handling missing data.

5. Running the analysis: Once all the variables and options have been selected, the user can run the CCA analysis.

6. Interpreting the results: The output of the CCA analysis will include various tables and charts, such as correlation coefficients and canonical loadings, which can be used to understand the relationship between the two data sets.

7. Checking assumptions: It is important to check for assumptions, such as linearity and normality, before interpreting the results of CCA.

8. Reporting the findings: The results of CCA should be reported in a clear and concise manner, using appropriate statistical language.

In conclusion, the procedure for conducting Canonical Correlation Analysis in SPSS involves data preparation, accessing the tool, selecting variables, setting options, running the analysis, interpreting the results, checking assumptions, and reporting the findings. It is a useful tool for exploring the relationship between two sets of variables and can provide valuable insights in various fields such as psychology, education, and marketing.

Canonical Correlation Analysis | SPSS Data Analysis Examples

Version info: Code for this page was tested in IBM SPSS 20.

Canonical correlation analysis is used to identify and measure the associations among two sets of variables. Canonical correlation is appropriate in the same situations where multiple regression would be, but where are there are multiple intercorrelated outcome variables. Canonical correlation analysis determines a set of canonical variates, orthogonal linear combinations of the variables within each set that best explain the variability both within and between sets.

Please Note: The purpose of this page is to show how to use various data analysis commands. It does not cover all aspects of the research process which researchers are expected to do. In particular, it does not cover data cleaning and checking, verification of assumptions, model diagnostics and potential follow-up analyses.

Examples of canonical correlation analysis

Example 1.

A researcher has collected data on three psychological variables, four academic variables (standardized test scores) and gender for 600 college freshman. She is interested in how the set of psychological variables relates to the academic variables and gender. In particular, the researcher is interested in how many dimensions are necessary to understand the association between the two sets of variables.

Example 2. A researcher is interested in exploring associations among factors from two multidimensional personality tests, the MMPI and the NEO. She is interested in what dimensions are common between the tests and how much variance is shared between them. She is specifically interested in finding whether the neuroticism dimension from the NEO can account for a substantial amount of shared variance between the two tests..

Description of the data

Let's pursue Example 1 from above.

We have included the data file, which can be obtained by clicking on mmreg.sav. The dataset has 600 observations on eight variables. The psychological variables are locus of control, self-concept and motivation. The academic variables are standardized tests in reading, writing, math and science. Additionally, the variable female is a zero-one indicator variable with the one indicating a female student.

Let's look at the data.

```
get file='d:datammreg.sav'.

descriptives
variables=locus_of_control self_concept motivation
read write math science female
/statistics=mean stddev min max.
```

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
locus of control	600	-2.23	1.36	.0965	.67028
self-concept	600	-2.62	1.19	.0049	.70551
motivation	600	.00	1.00	.6608	.34273
reading score	600	28.30	76.00	51.9018	10.10298
writing score	600	25.50	67.10	52.3848	9.72645
math score	600	31.80	75.50	51.8490	9.41474
science score	600	26.00	74.20	51.7633	9.70618
female	600	.00	1.00	.5450	.49839
Valid N (listwise)	600				

frequencies

Statistics

female

N	Valid	600
	Missing	0

female

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	273	45.5	45.5	45.5
	1.00	327	54.5	54.5	100.0
	Total	600	100.0	100.0	

variables=female .

Here are the correlations among the variables in the analysis.

correlations /variables=locus_of_control self_concept motivation read write math science female.

Correlations

		locus of control	self-concept	motivation	reading score	writing score	math score	science score	female
locus of control	Pearson Correlation	1	.171	.245	.374	.359	.337	.325	.113
	Sig. (2-tailed)		.000	.000	.000	.000	.000	.000	.005
	N	600	600	600	600	600	600	600	600
self-concept	Pearson Correlation	.171	1	.289	.061	.019	.054	.070	-.126
	Sig. (2-tailed)	.000		.000	.138	.634	.190	.087	.002
	N	600	600	600	600	600	600	600	600
motivation	Pearson Correlation	.245	.289	1	.211	.254	.195	.116	.098
	Sig. (2-tailed)	.000	.000		.000	.000	.000	.005	.016
	N	600	600	600	600	600	600	600	600
reading score	Pearson Correlation	.374	.061	.211	1	.629	.679	.691	-.042
	Sig. (2-tailed)	.000	.138	.000		.000	.000	.000	.307
	N	600	600	600	600	600	600	600	600
writing score	Pearson Correlation	.359	.019	.254	.629	1	.633	.569	.244
	Sig. (2-tailed)	.000	.634	.000	.000		.000	.000	.000
	N	600	600	600	600	600	600	600	600
math score	Pearson Correlation	.337	.054	.195	.679	.633	1	.650	-.048
	Sig. (2-tailed)	.000	.190	.000	.000	.000		.000	.238
	N	600	600	600	600	600	600	600	600
science score	Pearson Correlation	.325	.070	.116	.691	.569	.650	1	-.138
	Sig. (2-tailed)	.000	.087	.005	.000	.000	.000		.001
	N	600	600	600	600	600	600	600	600
female	Pearson Correlation	.113	-.126	.098	-.042	.244	-.048	-.138	1
	Sig. (2-tailed)	.005	.002	.016	.307	.000	.238	.001	
	N	600	600	600	600	600	600	600	600

Analysis methods you might consider

Below is a list of some analysis methods you may have encountered.

Some of the methods listed are quite reasonable while others have either

fallen out of favor or have limitations.

Canonical correlation analysis

SPSS performs canonical correlation using the manova command. Don't look for

manova in the point-and-click analysis menu, its not

there. The manova command is one of SPSS's hidden gems that is often overlooked. Used with the discrim option, manova will compute the canonical correlation analysis.

Due to the length of the output, we will be making comments in several places along the way.

manova locus_of_control self_concept motivation WITH read write math science female
/ discrim all alpha(1)
/ print=sig(eigen dim) .

The number of possible canonical variates, also known as canonical dimensions, is equal to the number of variables in the smaller set (the variables to the left of "WITH" in this example, called "DEPENDENT variables" in SPSS output). In our example, the first set has three variables and the second set has five (called "COVARIATES" in SPSS output). This leads to three possible canonical variates for each set,

which corresponds to the three columns for each set and three canonical correlation coefficients in the output. Canonical dimensions are latent variables that are analogous to factors obtained in factor analysis, except that canonical variates also maximize the correlation between the two sets of variables. In general, not all the canonical dimensions will be statistically significant. A significant dimension corresponds to a significant canonical correlation and vice versa.

The output below begins with an overall multivariate test of the entire model using four different multivariate criteria. This is followed by the three canonical correlations and the multivariate tests of each of the dimensions. These results show that the first two of the three canonical correlations are statistically significant at the .05 level.

- -

The default error term in MANOVA has been changed from WITHIN CELLS to WITHIN+RESIDUAL. Note that these are the same for all full factorial designs.

*** * * * * * * * * * * * * * * * * A n a l y s i s o f V a r i a n c e * * * * * * * * * * * * * * * * ***

600 cases accepted.
0 cases rejected because of out-of-range factor values.
0 cases rejected because of missing data.
1 non-empty cell.

1 design will be processed.

- -

*** * * * * * * * * * * * * * * * * * A n a l y s i s o f V a r i a n c e -- Design 1 * * * * * * * * * * * * * * * * ***

EFFECT .. WITHIN CELLS Regression
Multivariate Tests of Significance (S = 3, M = 1/2, N = 295)

Test Name Value Approx. F Hypoth. DF Error DF Sig. of F

Pillais .25425 11.00057 15.00 1782.00 .000
Hotellings .31430 12.37633 15.00 1772.00 .000
Wilks .75436 11.71573 15.00 1634.65 .000
Roys .21538

- -

Eigenvalues and Canonical Correlations

Root No. Eigenvalue Pct. Cum. Pct. Canon Cor. Sq. Cor

1 .27450 87.33628 87.33628 .46409 .21538
2 .02887 9.18537 96.52164 .16751 .02806
3 .01093 3.47836 100.00000 .10399 .01081

- -

Dimension Reduction Analysis

Roots Wilks L. F Hypoth. DF Error DF Sig. of F

1 TO 3 .75436 11.71573 15.00 1634.65 .000
2 TO 3 .96143 2.94446 8.00 1186.00 .003

3 TO 3 .98919 2.16461 3.00 594.00 .091

- -

Here we have the overall multivariate tests for dimensionality.

We also have the canonical correlations as well how much variance of the dependent variables is explained by the dimensions.
For this particular model there are three canonical dimensions of which only the first
two are statistically significant. The first test of dimensions tests whether all three
dimensions combined are significant (they are), the next test tests whether dimensions 2 and 3
combined are significant (they are). Finally, the last test tests whether dimension
3, by itself, is significant (it is not). Therefore dimensions 1 and 2 must each be
significant.

EFFECT .. WITHIN CELLS Regression (Cont.)

Univariate F-tests with (5,594) D. F.

Variable Sq. Mul. R Adj. R-sq. Hypoth. MS Error MS F Sig. of F

locus_of .18062 .17372 9.72160 .37123 26.18789 .000
self_con .01957 .01131 1.16669 .49212 2.37076 .038
motivati .07874 .07098 1.10799 .10913 10.15338 .000

- -

Raw canonical coefficients for DEPENDENT variables Function No.

Variable 1 2 3

locus_of 1.25383 -.62148 .66169
self_con -.35135 -1.18769 -.82672
motivati 1.26242 2.02726 -2.00023

- -

Standardized canonical coefficients for DEPENDENT variables
Function No.

Variable 1 2 3

locus_of .84042 -.41656 .44352
self_con -.24788 -.83793 -.58326
motivati .43267 .69480 -.68554

- -

Correlations between DEPENDENT and canonical variables
Function No.

Variable 1 2 3

locus_of .90405 -.38969 .17562
self_con .02084 -.70874 -.70516
motivati .56715 .35089 -.74513

- -

Variance in dependent variables explained by canonical variables

CAN. VAR. Pct Var DEP Cum Pct DEP Pct Var COV Cum Pct COV

1 37.97982 37.97982 8.17994 8.17994

2 25.90966 63.88948 .72701 8.90694

3 36.11052 100.00000 .39050 9.29745

- -

Raw canonical coefficients for COVARIATES

Function No.

COVARIATE 1 2 3

read .04462 -.00491 -.02138

write .03588 .04207 -.09131

math .02342 .00423 -.00940

science .00503 -.08516 .10984

female .63212 1.08464 1.79465

- -

Standardized canonical coefficients for COVARIATES

CAN. VAR.

COVARIATE 1 2 3

read .45080 -.04961 -.21601

write .34896 .40921 -.88810

math .22047 .03982 -.08848

science .04878 -.82660 1.06608

female .31504 .54057 .89443

- -

Correlations between COVARIATES and canonical variables

CAN. VAR.

Covariate 1 2 3

read .84045 -.35883 -.13536

write .87654 .06484 -.25456

math .76395 -.29795 -.14776

science .65841 -.67680 .23036

female .36411 .75493 .54340

- -

Variance in covariates explained by canonical variables

CAN. VAR. Pct Var DEP Cum Pct DEP Pct Var COV Cum Pct COV

1 11.30458 11.30458 52.48769 52.48769

2 .70132 12.00590 24.99409 77.48177

3 .09804 12.10394 9.06617 86.54795

- -

The raw canonical coefficients above are used to generate the canonical variates,
represented by the columns (1 2 3) in the coefficient tables,
for each set. They are interpreted in a manner analogous to interpreting
regression coefficients i.e., for the variable read, a one unit increase in reading leads to a
.0446 increase in the first canonical variate of the COVARIATE set when all of the
other variables are held constant. Here is another example: being female leads
to a .6321 increase in the dimension 1 for the COVARIATE set with the other predictors held constant.

The raw canonical coefficients are interpreted in a manner analogous to interpreting
regression coefficients i.e., for the variable read, a one

unit increase in reading leads to a .0446 increase in the first canonical variate of set 2 when all of the other variables are held constant. Here is another example: being female leads to a .6321 increase in the dimension 1 for set 2 with the other predictors held constant. When the variables in the model have very different standard deviations, the standardized coefficients allow for easier comparisons among the variables.

The raw canonical coefficients are followed by the standardized canonical coefficients. The standardized canonical coefficients are interpreted in a manner analogous to interpreting standardized regression coefficients. For example, consider the variable read, a one standard deviation increase in reading leads to a 0.45 standard deviation increase in the score on the first canonical variate for the COVARIATE set when the other variables in the model are held constant.

Things to consider

See also

References