Analysis of Multivariate Data

Canonical Correlation Analysis: Methodology and Application to Student Mental Well-Being

University of Cologne

Julian Bach

August 18, 2025

Table of Contents

1 Introduction

Canonical correlation analysis (CCA) was first discovered in 1935 by Hotelling (reprinted in Hotelling, 1992) and is one of the oldest multivariate techniques. CCA investigates the relationship between two different sets of quantitative variables measured across the same observations. The statistical method finds linear combinations of both sets (canonical variates) to maximize their correlation. The sets contain different type of variables e.g.: personality traits and ability measures, price indices and production indices or ecological variables and environmental variables (Rencher & Christensen, 2012). Therefore, CCA has applications across many fields such as Psychology, Economics and Ecology. Recent studies have tried to apply CCA to more than two sets of variables and developed cross validation approaches to improve stability towards unseen data (Abdi et al., 2018).

A convenient way to apply CCA is: find the canonical variates (see section 3.2), perform tests of significance (see section 3.3) and interpret the results (see section 3.4). The first objective of this paper is to provide an explanation of the method of CCA including significance tests. The second objective is to apply CCA to a dataset investigating students' life circumstances and mental well-being (Kaggle, 2024). The dataset is described within section 2. During the last decade, studies have detected a decline of students' mental well-being (Pietch, 2025; The Lancet Psychiatry, 2024). The Healthy Minds Study collected data of 373 campuses in the United States during 2013 and 2021 (Lipson et al., 2022). The study shows that in 2021 over 60% of students' fulfilled at least one criteria for mental health problems - an increase of over 50% from 2013. Therefore, examining the extent to which students' well-being is associated with their life circumstances is expected to provide valuable insights. If there is significant correlation, we are interested in the factors contributing the most to students' unhappiness (see section 4).

2 Problem

The original data set consists of 21 variables measured across 1238 students. We have created a subset with the variables of interest splitted into two sets: life circumstances and mental well-being. The variables included in each set are presented in table 1 and table 2 along with their descriptive statistics:

Table 1: life circumstances

Statistic	Min	Max	St. Dev.	Mean
Social_Support	2.24	10.00	1.47	6.93
Financial_Status	1.00	10.00	1.96	5.89
Work_Life_Balance	1.00	10.00	2.01	5.94
Freedom_to_Make_Life_Choices	1.00	10.00	1.94	5.94
Sports_Engagement	1.00	10.00	2.00	5.09
Average_Sleep_Hours	4.00	10.00	1.00	7.04
Anxiety	1.00	10.00	1.71	3.22
Isolation	1.00	9.23	1.71	3.13

Table 2: mental well-being

Statistic	Min	Max	St. Dev.	Mean
Mental_Health	1.25	10.00	1.52	7.00
Healthy_Life_Expectancy	5	10	1.69	7.52
Happiness_Level	1.00	7.84	1.00	4.81

We will perform CCA to identify the most influential factors within life circumstances affecting mental well-being, as well as the variables in mental well-being most impacted by these factors. By construction, the variables in one set are expected to share a high correlation which with CCA is not problematic but helpful.

Further information regarding the observation of the variables is unknown. We assume variables such as *happiness level*, *freedom to make life choices* etc. to be metric. However, this assumption may not hold as individuals have different internal scales. The findings should therefore be interpreted carefully as measurement errors cannot be ruled out.

3 Canonical Correlation Analysis (CCA)

This section is based on Rencher and Christensen (2012, pp. 361-373).

3.1 Mathematical Background

Let $x \in \mathbb{R}^q$ and $y \in \mathbb{R}^p$ be two sets of variables. If used, the index $j = 1, ..., q$ refers to j-th Variable in x and $k = 1, ..., p$ refers to the k-th variable in y. For instance, the sample covariance between x_k, y_j will be written as s_{kj}.

The portioned sample covariance matrix **S** contains of the covariance entries in and between x and y:

$$\mathbf{S} = \begin{pmatrix} S_{yy} & S_{yx} \\ S_{xy} & S_{xx} \end{pmatrix}$$

.

- $S_{yy} \in \mathbb{R}^{p \times p}$ contains p distinct covariance entries for y.
- $S_{xx} \in \mathbb{R}^{q \times q}$ contains q distinct covariance entries for x.
- $S_{xy} \in \mathbb{R}^{q \times p}$ and $S_{yx} \in \mathbb{R}^{p \times q}$ contains the covariance entries between x and y.

The portioned correlations matrix sums up the correlation structure in and between x and y. The dimensions of the entries in **R** are analogous to **S**. We define:

$$\mathbf{R} = \begin{pmatrix} R_{yy} & R_{yx} \\ R_{xy} & R_{xx} \end{pmatrix}$$

with $R_{xx} = D_{xx}^{-\frac{1}{2}} \mathbf{S} D_{xx}^{-\frac{1}{2}}, \quad R_{xy} = R_{yx}^T = D_{xy}^{-\frac{1}{2}} \mathbf{S} D_{xy}^{-\frac{1}{2}}, \quad R_{yy} = D_{yy}^{-\frac{1}{2}} \mathbf{S} D_{yy}^{-\frac{1}{2}}$
where $D_{yy} = \text{diag}(s_{y_1}^2, \ldots, s_{y_p}^2), \quad D_{xy} = \text{diag}(s_{x_1 y_1}, \ldots, s_{x_q y_p}), \quad D_{xx} = \text{diag}(s_{x_1}^2, \ldots, s_{x_q}^2).$

The coefficient of multiple determination R_M^2 is a measure for the (linear) relationship between y and x:

$$R_M^2 = |\underbrace{S_{yy}^{-1} S_{yx} S_{xx}^{-1} S_{xy}}_{M_{y|x}}| = |\underbrace{S_{xx}^{-1} S_{xy} S_{yy}^{-1} S_{yx}}_{M_{x|y}}|.$$

R_M^2 can be written in terms of it's (non-zero) eigenvalues $r_1^2, ...r_s^2$:

$$R_M^2 = |M_{y|x}| = |M_{x|y}| = \prod_{i=1}^{s} r_i^2 \tag{1}$$

with $s = \min(q, p)$. In (1), R_M^2 corresponds to the share of variation in y which is explained by the linear relationship to x. In the multivariate case, this is equivalent to the share of variation in x which is explained by the linear relationship to y. However, the product of the eigenvalues will be too small to contain any meaningful information about the correlation between x, y. The eigenvalues provide a good measure for the correlation between both sets and are important for the development of CCA.

3.2 The Principle

CCA aims to find linear combinations of x and y maximizing the correlation between both sets.

Definition 1. *The linear combinations of x and y are called canonical variates.*

$$u_i = \sum_j a_{ij} y_{ij}, \quad i = 1, ...s, j = 1, ..., p$$

$$v_i = \sum_k b_{ik} x_{ik}, \quad i = 1, ...s, k = 1, ..., q$$

Definition 2. *The canonical correlations correspond to the correlation of the canonical variates.*
The squared canonical correlations are identical to the eigenvalues of $M_{y|x}$ and $M_{x|y}$.

$$r_i = \max_{a_i, b_i} r_{u_i, v_i}$$

In general, there are $s = \min(q, p)$ canonical correlations with corresponding canonical variates. However, the first canonical correlation is the maximum correlation between both sets. The computation of the canonical correlations and canonical variates is based on (1).

1. **Compute the eigenvalues**

$$|M_{y|x} - r^2 I| = 0 \tag{2}$$

$$|M_{x|y} - r^2 I| = 0 \tag{3}$$

   Solving equation (2) and (3) will lead to the same eigenvalues (see (1)).

   The s eigenvalues are ordered respectively, that is: $r_1^2 \geq r_2^2 ... \geq r_s^2$.

2. **Compute the coefficient vectors or eigenvectors** a_i, b_i

$$\left(M_{y|x} - r_i^2 I\right) a_i = 0 \tag{4}$$

$$\left(M_{x|y} - r_i^2\right) b_i = 0 \tag{5}$$

An example is calculated in A.1. The canonical correlations and canonical variates are computed using the CCA software package by González et al. (2008). In most studies, p is smaller than q. As rank$(M_{y|x}) = p$, solving (2) and (3) will lead to p non-zero eigenvalues and $(q-p)$ eigenvalues corresponding to zero. Therefore, there are $s = \min(q, p) = p$ canonical correlations with corresponding canonical variates. The canonical variates $u_1, ..., u_s$ and $v_1, ..., v_s$ are uncorrelated with each other:
$r_{u_j, v_k} = r_{u_j, u_k} = r_{v_j, v_k} = 0$ for all $j, k = 1, ..., s$ and $j{\neq}k$.

Not all canonical correlations provide useful information regarding the correlation between both sets. The relative importance of r_i can be judged by:

$$r_i^- = \frac{r_i^2}{\sum_{j \neq i} r_j^2}$$

.

Instead of the method above, we can obtain the canonical correlations and corresponding variates by using the portioned correlation Matrix. $M_{y|x}$ is replaced by $R_{yy}^{-1} R_{yx} R_{xx}^{-1} R_{xy}$ and $M_{x|y}$ is replaced by $R_{xx}^{-1} R_{xy} R_{yy}^{-1} R_{yx}$. The eigenvalues are identical. The coefficient vectors are now standardized which is beneficial for the interpretation (see section 3.4).

3.3 Tests of Significance

We present four possibilities to test the significance of the canonical correlations. Each is contained in the later used software package by Menzel (2020).

Testing for independence

If x and y are independent, they share no relationship and the canonical correlations are insignificant. We have:

$$H_0 : \mathbf{\Sigma} = \begin{pmatrix} \Sigma_{yy} & 0 \\ 0 & \Sigma_{xx} \end{pmatrix} = \mathbf{0} \quad \text{vs.} \quad H_1 : \mathbf{\Sigma} \neq \mathbf{0}.$$

Or in short: $\Sigma_{yx}, \Sigma_{xy} = \mathbf{0}$ as there are no restrictions regarding the correlation structure in one set. Testing for independence means testing if the correlations entries in Σ_{yx}, Σ_{xy} are (jointly) significantly different from zero. Wilk's lambda test statistic is given by:

$$\Lambda_1 = \frac{|\boldsymbol{S}|}{|S_{yy}||S_{xx}|} = \frac{|\boldsymbol{R}|}{|R_{yy}||R_{xx}|} \sim \Lambda_{p,q,n-1-q} \tag{6}$$

with p number of variables in y, q number of variables in x and $(n - 1 - q)$ degrees of freedom (d.o.f). H_0 is rejected at α% , if $\Lambda_1 < \Lambda_\alpha$ where Λ_α corresponds to the α% quantile for Wilk's lambda test statistic. The critical values for $\alpha = .05$% can be found in Rencher and Christensen (2012, pp. 566–573). Under H_0, $|\boldsymbol{S}| \approx |S_{xx}||S_{yy}|$ as there is almost no correlation between x, y. Therefore, the smaller Λ_1, the more correlation between both sets. It holds that $|\boldsymbol{S}| < |S_{xx}||S_{yy}|$ (Rencher & Christensen, 2012, p. 260).

Alternatively, we can construct the test based on the eigenvalues of $M_{y|x}$:

$$\Lambda_1 = \prod_{i=1}^{s}(1 - r_i^2) \sim \Lambda_{q,p,n-1-p} \tag{7}$$

with $s = \min(q, p), p, q$ and d.o.f. defined as in (6). If H_0 is rejected taking r_1 into account, there is no information regarding the significance of the succeeding canonical correlations. The test for the k-th canonical correlation is performed using (7), removing $r_1, ...r_{k-1}$ and adjusting p, q and $(n - 1 - p)$ d.o.f. accordingly (see A.2).

Often the range of parameters exceeds the range of critical values for Λ. By transforming Λ_1, we can test significance of the canonical correlations with a Fisher approximation:[1]

$$\Lambda_1 \approx \frac{1-\Lambda_1^{\frac{1}{t}}}{\Lambda_1^{\frac{1}{t}}} \cdot \frac{df_2}{df_1} \sim F_{df_2,df_1} \tag{8}$$

with $df_2 = pq$,

$df_1 = t[n - \frac{1}{2}(p+q+3)] - \frac{1}{2}pq + 1$ and $t = \begin{cases} 1 & ; pq = 2 \\ \sqrt{\frac{(pq)^2-4}{p^2+q^2-5}} & ; \text{otherwise.} \end{cases}$

Testing canonical correlations

The canonical correlations can be tested directly for significance.[2] The presented test statistics follow the same idea: the larger $r_1, ..., r_s$, the more likely the null is about to be rejected. If the test result is significant, (at least) r_1 is significantly different from zero. If the test result is insignificant, the s canonical correlations are not significantly different from zero. We have:

$$H_0 : r_1, ...r_s = 0 \quad \text{vs.} \quad H_1 : r_1, ...r_s \neq 0 \text{ (for at least } r_1).$$

Again, approximations of the test statistics are needed if the range of parameters exceeds the range of critical values. For Fisher approximations of the following tests, see Rencher and Christensen (2012, chapter 6)

1. Pillai´s test statistic

$$V^{(s)} = \sum_{i=1}^{s} r_i^2 \tag{9}$$

   For large values of $V^{(s)}$, the null is rejected. Upper percentage points for $\alpha = .05\%$ are given in Rencher and Christensen (2012, pp. 578–581) with $s = \min(q,p), m = \frac{1}{2}(|q-p|-1)$ and $N = \frac{1}{2}(n-q-p-2)$. For testing the significance of the k-th canonical correlation, $r_1, ...r_{k-1}$ are removed and s, m and N adjusted respectively.

[1] A Chi-square approximation is applicable but not further addressed in this context. For more details, see Rencher and Christensen (2012, p. 367)). For $s = \{1, 2\}$, the test statistic in (8) follows an exact F-distribution and the Fisher approximation is preferred over the Chi-square approximation.

[2] Pillai´s test statistic (9) and Lawley's Hotelling test statistic (10) are defined differently throughout the book. Here, we adhere to the definition established within the framework of CCA.

2. Lawley Hotelling statistic

$$U^{(s)} = \prod_{i=1}^{s} \frac{r_i^2}{1 - r_i^2} \tag{10}$$

Again, the null is rejected for large values of $U^{(s)}$. Upper percentage points for $\alpha =$ .05% can be found in Rencher and Christensen (2012, pp. 582–586) with $v_E = n-q-1$ and $v_h = q$. Testing the k-th canonical correlation is analogous as discussed in (9) with v_E, v_H and s being adjusted respectively.

3. Roy´s largest root statistic

$$\theta = r_1^2 \tag{11}$$

The null is rejected for large values of θ. Upper percentage points for $\alpha =$ .05% can be found in Rencher and Christensen (2012, pp. 574–577) with $s = \min(q, p)$, $m = \frac{1}{2}(|q-p| - 1)$ and $N = \frac{1}{2}(n - q - p - 2)$. Note that Roy´s test statistic only tests the significance of the first canonical correlation.

3.4 Interpretation of CCA

Property 1. *Canonical correlations are scale invariant, eigenvectors are scale variant.*[3]

Property 2. *The first canonical correlation corresponds to the maximum correlations between any linear function regarding x and y. This holds for for all possible subsets of both sets.*

The coefficient vectors a_i, b_i contain information about the contribution of each variable to the correlation between x and y. By property 1, the eigenvectors are not scale invariant and therefore have to be standardized to control for measuring differences. If a_i and b_i are computed with the portioned correlation matrix, they are standardized (see section 3.2). If not, they have to be transformed using:

[3] For instance: if x and y are originally measured in dollar but now transformed to euros, the canonical correlations are identical. The eigenvectors are not.

$$c_i = D_y a_i \qquad d_i = D_x b_i \tag{12}$$

with $D_y = \text{diag}(y_1, \ldots, y_p)$ and $D_x = \text{diag}(x_1, \ldots, x_q)$.

After being transformed, the entries in $c_i^T = (c_1, \ldots, c_p)$ correspond to the relative contribution of $y_i^T = (y_1, \ldots \ , y_p)$ on u_i. The interpretation for $d_i^T = (d_1, \ldots, d_p),\ x_i^T = (x_1, \ldots, x_q)$ and v_i is analogous.

The impact channels for the i-th canonical correlation r_i are summarized in figure 1. For simplicity, let $y^T = (y_1, y_2)$ and $x^T = (x_1, x_2)$ and therefore $s = i = 1, 2$.

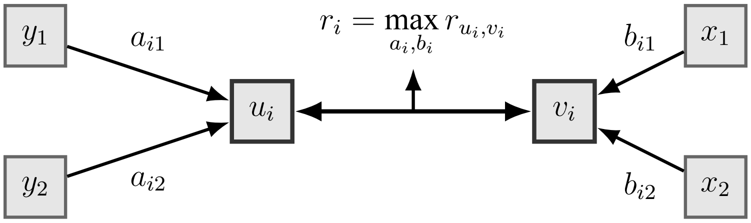

Figure 1: impact channels of canonical correlations

4 Empirical Analysis

For a descriptive analysis and obtaining the canonical variates, we will use the CCA software package developed by González et al. (2008). The variables in life circumstances correspond to x and the variables in mental well-being correspond to y. Before applying CCA, visualizing the correlation matrices is helpful (see figure 2).

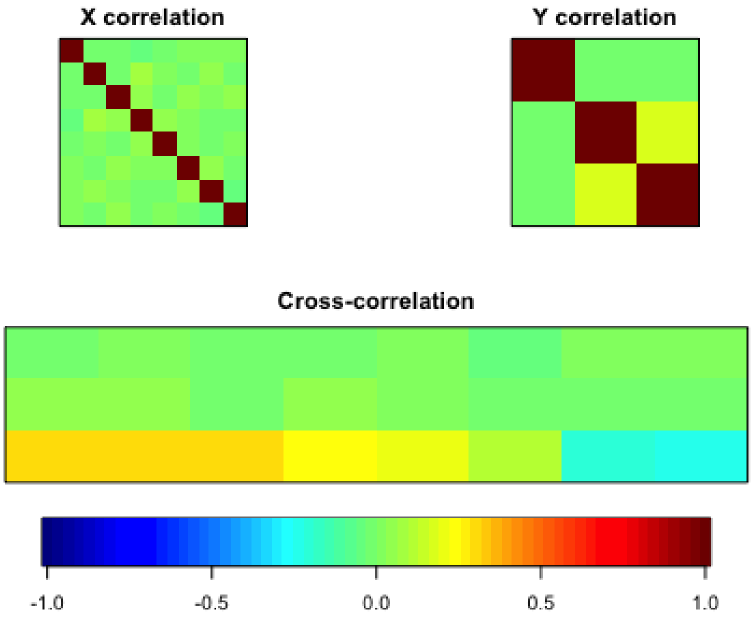

Figure 2: correlation matrices

The correlation between the variables in x and y are shown above, with color coding according to the given legend. Contrary to our initial assumptions, the correlation between the variables in x, y is low. The cross-correlation between x,y - that is the correlation between life circumstances and mental well-being - does not indicate significance either expect at the bottom.[4] The lower part of the cross-correlation matrix represents the correlation between *happiness levels* and life circumstances.

Following section 3.2, we now obtain the canonical correlations and their corresponding canonical variates. Ordering the three canonical correlations respectively we have:

$$r_1 = 0.698 > r_2 = 0.066 > r_3 = 0.050.$$

As discussed within the section, not all canonical correlations provide useful information regarding the correlation between x and y. Here, only the first canonical correlation is worth considering as a dimension of linear relationship. The corresponding canonical variates for r_1, r_2, r_3 are denoted in table 3 and table 4:[5]

[4] According to González et al. (2008) the work must be stopped if the visualization of the cross correlation matrix mostly shows green colors. Here we continue.

[5] A visualization of the canonical variates is useful if at least more than one canonical correlation is significantly different from zero. For more details, see González et al. (2008, pp. 11–12.)

Table 3: canonical variates life circumstances

Statistic	r_1	r_2	r_3
Social_Support	0.462	0.313	-0.087
Financial_Status	0.444	0.330	0.353
Work_Life_Balance	0.460	-0.635	0.495
Freedom_to_Make_Life_Choices	0.273	0.280	-0.456
Sports_Engagement	0.280	0.294	0.180
Average_Sleep_Hours	0.141	-0.407	-0.489
Anxiety	-0.343	0.185	0.083
Isolation	-0.357	0.129	0.415

Table 4: canonical variates mental well-being

Statistic	r_1	r_2	r_3
Mental_Health	-0.014	0.399	0.917
Healthy_Life_Expectancy	-0.130	0.933	-0.387
Happiness_Level	1.016	-0.044	0.052

Before interpreting the canonical variates, we perform asymptotic tests of significance as outlined in section 3.3. For this purpose, we use the CCP software package developed by Menzel (2020). Given the large number of observations, the assumption of asymptotic multivariate normality for x and y is not a concern. As discussed in section 3.4, there are multiple ways to access significance. Using a Fisher-approximation of Wilk's lambda test statistic or testing for independence as discussed in (8), we show that only the first canonical correlation is significantly different from zero. The output is summarized in table 5:

Table 5: significance test for independence

	id	stat	approx	df1	df2	p.value
1	Wilks	0.51	38.95	24.00	3562.17	0.00
2	Wilks	0.99	0.60	14.00	2458.00	0.87
3	Wilks	1.00	0.52	6.00	1230.00	0.79

We obtain the same result using any other mentioned test statistic within section 3.3.[6] Therefore, we only consider the canonical variates for r_1. For life circumstances, *social support*, *financial status* and *work–life-balance* show the most positive contribitution to mental well-being whereas *anxiety* and *isolation* contribute negatively. For mental well-being, only *happiness level* makes a relevant contribution to the correlation. Since the variables in mental well-being share little correleation (see figure 2), CCA limits the relationship to the variable the most correlated with life circumstances - *happiness level*. As discussed in section 2, measurement errors may reduce the correlation between variabels in mental well-being.

5 Conclusion

We have presented a methodological overview of CCA including the computation of the canonical correlations and canonical variates, performing tests of significance and the interpretation of the results. We then applied the method of CCA to a data set focusing on students' life circumstances and their mental well-being. The variables in both sets share little correlation for which measurement errors could be responsible. After performing asymptotic tests, only the first canonical correlation remained significant. To counter the decline of students' mental well-being, we found that *social support*, *financial status* and *work life balance* have the most positive impact on student mental well-being whereas *anxiety* and *isolation* have a negative impact. For mental well-being, CCA effectively only considers *happiness level*. Further research could employ non-parametric CCA or investigate new data to find additional dimensions of relationship extracting more information related to mental well-being.

[6] Note that Roy's test statistic only tests the significance of the first canonical correlation. Therefore, the outcome only shows that the first canonical correlation is significant.

A Appendix

A.1 Calculation of the Canonical Correlations and Canonical Variates

Let $x, y \in \mathbb{R}^2$. Therefore, the number of non-zero eigenvalues is $s = \min(2,2) = 2$. The eigenvalues have to be computed first by (2) and (3).

$$|M_{y|x} - \lambda_1 I| = 0 \iff \left| \begin{pmatrix} m_1 - \lambda & m_2 \\ m_3 & m_4 - \lambda \end{pmatrix} \right| = 0$$

$$\iff (m_1 - \lambda)(m_4 - \lambda) - m_2 m_3 = 0 \iff \lambda^2 + m_1 m_4 - \lambda(m_1 + m_4) - m_2 m_3 = 0$$

Solving the quadratic equation leads to two eigenvalues λ_1, λ_2 or squared canonical correlations. The squared canonical correlations are ordered respectively: $\lambda_1 = r_1^2 \geq \lambda_2 = r_2^2$. The first canonical variate is obtained by inserting λ_1 and solving both (4) and (5).

$$I : \left(\begin{pmatrix} m_1 & m_2 \\ m_3 & m_4 \end{pmatrix} - \lambda_1 I \cdot \right) a_i = 0 \iff \begin{pmatrix} m - \lambda_1 & m_2 \\ m_3 & m_4 - \lambda_1 \end{pmatrix} \cdot \begin{pmatrix} a_{11} \\ a_{12} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$II : \left(\begin{pmatrix} m_1 & m_2 \\ m_3 & m_4 \end{pmatrix} - \lambda_2 I \cdot \right) b_1 = 0 \iff \begin{pmatrix} m - \lambda_2 & m_2 \\ m_3 & m_4 - \lambda_2 \end{pmatrix} \cdot \begin{pmatrix} b_{11} \\ b_{12} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

The solution for a_1, b_1 corresponds to the first canonical coefficients $a_1 = (a_{11}, a_{12})$, $b_1 = (b_{11}, b_{12})$. Therefore, the first canonical variates are $u_1 = a_1^T y$ and $v_1 = b_1^T x$. The second canonical variate is computed analogously by inserting λ_2.

A.2 Testing the k-th Canonical Correlation

The k-th canonical correlation is tested using (7) and adjusting p, q and d.o.f.
accordingly:

$$\Lambda_k = \prod_{i=k}^{s} \sim \Lambda_{p-k+1, q-k+1, n-k-1}.$$

The null is rejected, if $\Lambda_k < \Lambda_\alpha$. If the range of parameters exceeds the range of critical values, a Fisher approximation is needed. The test statistic is given by:

$$\Lambda_k \approx \frac{1-\Lambda_k^{\frac{1}{t}}}{\Lambda_k^{\frac{1}{t}}} \cdot \frac{df_2}{df_1} \sim F_{df_2, df_1}$$

with $df_2 = (p-k+1)(q-k+1)$,
$df_1 = t[n - \frac{1}{2}((p-k+1)+(q-k+1)+3)] - \frac{1}{2}(p-k+1)(q-k+1)+1$
and $t = \sqrt{\frac{(p-k+1)^2(q-k+1)^2-4}{(p-k+1)^2+(q-k+1)^2-5}}$.

References

Abdi, H., Guillemot, V., Eslami, A., & Beaton, D. (2018). History of CCA. In R. Alhajj & J. Rokne (Eds.), *Encyclopedia of social network analysis and mining*. Springer. https://doi.org/10.1007/978-1-4939-7131-2_110191

González, I., Déjean, S., Martin, P., & Baccini, A. (2008). CCA: An R package to extend canonical correlation analysis. *Journal of Statistical Software*, *23*(12), 1–14. https://doi.org/10.18637/jss.v023.i12

Hotelling, H. (1992). *Relations between two sets of variates* (S. Kotz & N. L. Johnson, Eds.). https://doi.org/10.1007/978-1-4612-4380-9_14

Kaggle. (2024). *Student's mental well-being* [Retrieved August 18, 2025]. https://www.kaggle.com/datasets/israelfiyinfoluwa/student-mental-well-being

Lipson, S. K., Zhou, S., Abelson, S., Heinze, J., Jirsa, M., Morigney, J., Patterson, A., Singh, M., & Eisenberg, D. (2022). Trends in college student mental health and help-seeking by race/ethnicity: Findings from the national healthy minds study, 2013–2021. *Journal of Affective Disorders*, *306*, 138–147. https://doi.org/https://doi.org/10.1016/j.jad.2022.03.038

Marek, H. (2022). *Stargazer*. https://cran.r-project.org/web/packages/stargazer/vignettes/stargazer.pdf

Menzel, U. (2020). *Significance tests for canonical correlation analysis (CCA)*. https://cran.r-project.org/web/packages/CCP/CCP.pdf

Pietch, F. (2025). *Mental health barometer* [Retrieved August 18, 2025]. https://studo.com/de/blog/mental-health-barometer-2024

Rencher, A. C., & Christensen, W. F. (2012). *Methods of multivariate analysis*. John Wiley & Sons. https://doi.org/10.1002/9781118391686

The Lancet Psychiatry. (2024). Prioritising young people. *The Lancet Psychiatry*, *11*(9), 665. https://www.thelancet.com/journals/lanpsy/article/PIIS2215-0366(24)00252-9/fulltext