Canonical Correlation Analysis of the Relationship between Personality Traits and Cognitive Ability

N.W.I.M.S.Sewwandi
(S/18/511)

Mini Project 02 ST405

Department of Statistics and Computer Science
Faculty of Science
University of Peradeniya
2024

1 Introduction

Personal traits and cognitive ability are some elemental constructs of psychology. Those can be use to predict one's psychological preparedness for his or her career because those might be differ from job to job (Mullola et al., 2018). Thus, investigating relationship between personal traits and cognitive ability is very important. Also, it can be helps improve matters such as personal well-being and productivity. The major purpose of the study is to check whether personal traits and cognitive ability are related. Personalty traits are significantly correlated with cognitive ability is the hypothesis that is going to be tested in this analysis.

Personality is a combination of characteristics that causes consistent patterns of behavior, thoughts, and feelings in an individual (Diener & Lucas, 2019). The OCEAN (Openness, Conscientiousness, Extroversion, Agreeableness, Neuroticism) Model of Personality is the most popular framework to describe human personality traits (Durupinar, Pelechano, Allbeck, Güdükbay, & Badler, 2009). Cognitive ability reflects n one's ability to understand information, learn, and use new ideas and knowledge to solve problems (Cherry, 2024). Numerical aptitude, Spatial aptitude, Perceptual Aptitude, Abstract reasoning, Verbal reasoning are used to test cognitive ability.

2 Methodology

The data set is collected from Kaggle website. It consists with 11 variables and 105 individuals. Table 1 shows the descriptive details of variables. All variables from 1 to 10 represent scores given for each measurement. Last variables gives the job title.

> **Source:** Shrivastav, U. (2024, March). Career Prediction Dataset. Retrieved May 20, 2024, from https://www.kaggle.com/datasets/utkarshshrivastav07/career-prediction-dataset/data

Canonical Correlation Analysis is used to investigate the relationship between personal traits and cognitive ability.

Table 1: Description of Variables

Index	**Variable**	**Data type**	**Description**
1	O_score	Numerical	Score for Openness
2	C_score	Numerical	Score for Conscientiousness
3	E_score	Numerical	Score for Extroversion
4	A_score	Numerical	Score for Agreeableness
5	N_score	Numerical	Score for Neuroticism
6	Numerical Aptitude	Numerical	Score for proficiency in numerical reasoning
7	Spatial Aptitude	Numerical	Score for ability to mentally manipulate shapes, visualize objects in different orientations
8	Perceptual Aptitude	Numerical	Score for ability to accurately perceive and interpret information of relevant field
9	Abstract Reasoning	Numerical	Score for potential success in certain fields.
10	Verbal Reasoning	Numerical	Score for ability to comprehend and analyze written information, draw logical conclusions, and solve problems based on textual data.
11	Career	Categorical	Job title

The variables were divided into two sets are as follows.

- Variables from 1 to 5 = Set A ($\mathbf{U}$) : Personality traits
- Variables from 6 to 10 = Set B ($\mathbf{V}$) : Cognitive ability

Preliminary Analysis
Data cleaning was done. A new dataset without career variable (Table 1 1) was obtained. Descriptive statistics were obtained for all variables. Normality of data was tested using Shapiro-Wilk Test (Hernandez, 2021). Then, the correlations between variables of A, variables of B, and variables of A & B variables were calculated using Pearson method.
Canonical Correlation Analysis
Canonical correlations and correlations between variables and the canonical variates were obtained. Significance of canonical correlations are tested using Wilks' lambda test and Lawley-Hotelling test. **tidyverse, GGally, CCA, CCP** were the packages of R were used in analysis (González, Déjean, Martin, & Baccini, 2008). The i th canonical pair is represented by $(U_i, V_i) : i = (1, 2, 3, 4, 5)$.

3 Results and Discussion

O_score	C_score	E_score	A_score	N_score	Numerical Aptitude	Spatial Aptitude	Perceptual Aptitude	Abstract Reasoning	Verbal Reasoning
Min. :2.670	Min. :3.560	Min. :2.890	Min. :3.230	Min. :2.890	Min. :2.89	Min. :2.340	Min. :3.010	Min. :3.010	Min. :3.450
1st Qu.:6.670	1st Qu.:7.340	1st Qu.:4.230	1st Qu.:5.450	1st Qu.:4.670	1st Qu.:4.45	1st Qu.:3.120	1st Qu.:3.670	1st Qu.:4.340	1st Qu.:5.450
Median :7.230	Median :7.670	Median :5.230	Median :6.450	Median :5.450	Median :5.12	Median :3.450	Median :4.450	Median :4.670	Median :7.450
Mean :7.295	Mean :7.538	Mean :5.549	Mean :6.864	Mean :5.466	Mean :5.94	Mean :4.376	Mean :5.164	Mean :5.724	Mean :6.794
3rd Qu.:8.670	3rd Qu.:8.340	3rd Qu.:7.010	3rd Qu.:8.120	3rd Qu.:6.010	3rd Qu.:7.78	3rd Qu.:4.450	3rd Qu.:6.780	3rd Qu.:7.670	3rd Qu.:8.120
Max. :9.450	Max. :9.450	Max. :9.340	Max. :9.340	Max. :8.120	Max. :9.45	Max. :9.230	Max. :9.340	Max. :9.340	Max. :9.340

Figure 1: Summary output of the dataset

No any empty values or duplicated data was found. According to figure 1, all scores are in between 2.34 and 9.45. O_score, C_score, A_score, Numerical Aptitude, Abstract Reasoning, Verbal Reasoning have higher means than Spatial Aptitude, Perceptual Aptitude, E_score, N_score.

Table 2: Normality Testing of Variables using Shapiro-Wilk Test

Varible	**Test Statistic**	**P-value**
O_score	0.9325196	$\ll 0.0001$
C_score	0.8173158	$\ll 0.0001$
E_score	0.9109629	$\ll 0.0001$
A_score	0.9404697	0.000140
N_score	0.9429652	0.000202
Numerical Aptitude	0.890694	$\ll 0.0001$
Spatial Aptitude	0.7295321	$\ll 0.0001$
Perceptual Aptitude	0.8740709	$\ll 0.0001$
Abstract Reasoning	0.8808683	$\ll 0.0001$
Verbal Reasoning	0.9251699	$\ll 0.0001$

According to P-values of Table 2, all data follows normality assumption (taking $\alpha = 0.05$). Hence, the correlations and scatter plots between variables were obtained (refer figure 2, figure 3, Table 3)

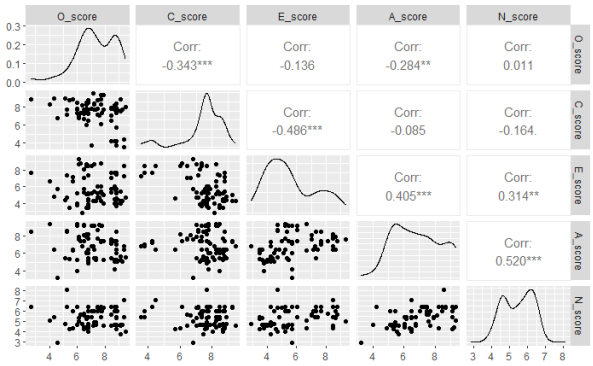

Figure 2: Correlations and Scatter plots of variables in Set A

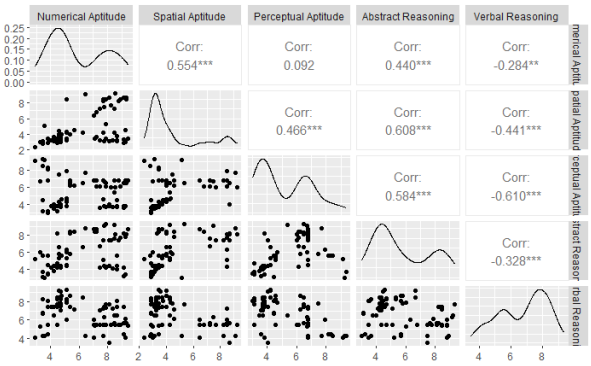

Figure 3: Correlations and Scatter plots of variables in Set B

Table 3: Correlations between Set A and Set B variables

	Numerical Ap:	**Spatial Ap:**	**Perceptual Ap:**	**Abstract Re:**	**Verbal Re:**
O_score	-0.09689467	0.01248246	0.35374284	0.39902332	-0.04911499
C_score	0.48845504	0.06116467	-0.53847746	0.03546569	0.39140913
E_score	-0.63681772	-0.38397130	-0.02268739	-0.39188283	-0.02909311
A_score	-0.63721953	-0.58718447	-0.42758847	-0.68201654	0.49186104
N_score	-0.64159361	-0.54972665	-0.28418245	-0.38699872	0.38249140

Considering set A, most of them have negative correlations while only between E_ score & C_score and N_ score & A_score have much moderate correlations (figure 2). For set B, most of them have positive correlations while only between Spatial Aptitude & Numerical Aptitude , Abstract Reasoning & Spatial Aptitude and Abstract Reasoning & Perceptual Aptitude have much moderate correlations (figure 3). Considering Table 3, most of them are negative correlations while E_score & Numerical Aptitude, A_score & Numerical Aptitude, N_score & Numerical Aptitude, C_score & Perceptual Aptitude,and A_score & Abstract Reasoning have comparatively moderate correlations. Because of reasonable correlations observed between sets, canonical correlation analysis was carried out.

Table 4: Canonical Correlations obtained

	Canonical Correlations	**Squared Canonical Correlations**
1	0.89410476	0.9455711
2	0.79596533	0.8921689
3	0.56700793	0.7529993
4	0.31353817	0.5599448
5	0.01427121	0.1194622

94.55% of the variation in U_1 is explained by the variation in V_1, 89.21%, 75.29%, 55.994% and 11.945% for other canonical variate pairs respectively (refer Table 4). Hence, that implies that may only 1st four correlations are important.

Significance of Canonical Correlations

According to figures 4 & 5 with taking $\alpha = 0.05$, this conforms that only from 1st to 4th canonical variate pairs are significantly correlated and depend on one another while 5th is not significant.

Wilks' Lambda, using F-approximation (Rao's F):

	stat	approx	df1	df2	p.value
1 to 5:	0.0449577	18.49891413	25	354.4113	0.000000e+00
2 to 5:	0.2241422	11.60102520	16	293.9225	0.000000e+00
3 to 5:	0.6116765	5.87450855	9	236.2231	2.146960e-07
4 to 5:	0.9015102	2.60725566	4	196.0000	3.699609e-02
5 to 5:	0.9997963	0.02016718	1	99.0000	8.873598e-01

Figure 4: Output of Wilks' lambda test

Hotelling-Lawley Trace, using F-approximation:

	stat	approx	df1	df2	p.value
1 to 5:	6.2976527193	23.52803056	25	467	0.000000e+00
2 to 5:	2.3120283810	13.78546922	16	477	0.000000e+00
3 to 5:	0.5830626220	6.31003326	9	487	1.808813e-08
4 to 5:	0.1092276188	2.71430633	4	497	2.937457e-02
5 to 5:	0.0002037089	0.02065608	1	507	8.857769e-01

Figure 5: Output of Lawley-Hotelling test

Considering the estimated canonical coefficients, canonical variables for Personality traits are
$U_1 = -0.0942O\_score + 0.1679C\_score - 0.2031E\_score - 0.3445A\_score - 0.3672N\_score$,
$U_2 = 0.1423O\_score + 0.8274C\_score + 0.0290E\_score + 0.2750A\_score + 0.2192N\_score$,
$U_3 = -0.6859O\_score - 0.3201C\_score - 0.1269E\_score0.2076A\_score - 0.4540N\_score$,
$U_4 = -0.2183\_score + 0.3963C\_score + 0.5811E\_score - 0.6268A\_score0.5000N\_score$.
Canonical variables for cognitive ability are
$V_1 = 0.3906Numerical + 0.0820Spatial - 0.0733Perceptual + 0.1424Abstract - 0.0128Verbal$,
$V_2 = 0.13054Numerical - 0.0469Spatial - 0.3658Perceptual + 0.0851Abstract + 0.3504Verbal$,
$V_3 = 0.1555Numerical + 0.3257Spatial + 0.0227Perceptual - 0.6787Abstract - 0.1298Verbal$,
$V_4 = -0.2731Numerical - 0.0450Spatial - 0.8059Perceptual + 0.3820Abstract - 0.8131Verbal$.

Figure 6 indicates U_1 shows strong relationships with E_score, A_score, N_score positively while with Numerical aptitude negatively. U_2 shows strong relationships with C_score positively and with Verbal Reasoning positively. U_3 shows strong relationships with O_score negatively. U_4 shows moderate relationship with E_score. V_1 performs strong relationship with Numerical aptitude positively and with A_score negatively. V_2 performs strong relationship with Verbal Reasoning and Perceptual Aptitude. V_3 performs moderate relationships with O_score and Abstract Reasoning. V_4 performs moderate relationship Verbal Reasoning. Since U_1 contains comparatively larger loadings, it can be taken as an overall measure of personality traits for cognitive ability, it is V_1.

4 Conclusion and Recommendation

The canonical correlation analysis concludes since, the first four canonical correlations between set A and set B were significant, it indicates there is a meaningful relationship between personality traits and cognitive ability. The first pair of canonical variates explains a significant portion of variance in both sets, while other pairs contribute lesser. Therefore, there are multiple dimensions that connects personality traits to cognitive ability. Since the sample size is 105, it is recommended to collect a higher number of samples using additional personality traits and aptitudes not covered in this study. This study can be expanded to investigate the applicability of findings across different populations for instances, people with different ages, geographical locations.

```
$corr.X.xscores
               [,1]       [,2]        [,3]       [,4]        [,5]
O_score -0.01123134 -0.2690361 -0.89111350 -0.3259500 -0.16482337
C_score  0.50818440  0.8470835  0.07978514  0.1311568 -0.02510856
E_score -0.72659057 -0.2398422  0.11494847  0.5143186 -0.36987169
A_score -0.80301948  0.3933888  0.33108318 -0.2918795 -0.07480144
N_score -0.73023366  0.2596170 -0.24327648  0.1583706  0.56133154

$corr.Y.xscores
                         [,1]       [,2]        [,3]          [,4]          [,5]
Numerical Aptitude   0.8624269  0.05341582  0.08833930 -0.0166407988 -0.0027823504
Spatial Aptitude     0.6210294 -0.30390934  0.07638971 -0.0219537373  0.0084275238
Perceptual Aptitude  0.1651048 -0.70438703 -0.14760958 -0.1061519680  0.0002759776
Abstract Reasoning   0.5651205 -0.26171297 -0.37681246  0.0008977763  0.0032149626
Verbal Reasoning    -0.2828948  0.66023895 -0.09515400 -0.1254897948  0.0021876199

$corr.X.yscores
               [,1]       [,2]        [,3]       [,4]         [,5]
O_score -0.01004199 -0.2141434 -0.50526843 -0.10219777 -0.0023522285
C_score  0.45437009  0.6742491  0.04523881  0.04112266 -0.0003583294
E_score -0.64964809 -0.1909060  0.06517669  0.16125853 -0.0052785157
A_score -0.71798353  0.3131239  0.18772679 -0.09151535 -0.0010675069
N_score -0.65290539  0.2066462 -0.13793969  0.04965524  0.0080108790

$corr.Y.yscores
                         [,1]       [,2]       [,3]         [,4]        [,5]
Numerical Aptitude   0.9645703  0.06710823  0.1557990 -0.053074235 -0.19496250
Spatial Aptitude     0.6945824 -0.38181228  0.1347242 -0.070019345  0.59052632
Perceptual Aptitude  0.1846593 -0.88494688 -0.2603307 -0.338561545  0.01933807
Abstract Reasoning   0.6320518 -0.32879946 -0.6645629  0.002863372  0.22527614
Verbal Reasoning    -0.3164001  0.82948204 -0.1678178 -0.400237694  0.15328905
```

Figure 6: Correlations between original variables and canonical variates (loadings)

References

Cherry, K. (2024, April). *What does 'cognitive' mean in psychology?* https://www.verywellmind.com/what-is-cognition-2794982#:~:text=In%20psychology%2C%20the%20term%20'cognitive,parts%20of%20the%20human%20experience.

Diener, E., & Lucas, R. E. (2019). Personality traits. *General psychology: Required reading*, *278*.

Durupinar, F., Pelechano, N., Allbeck, J., Güdükbay, U., & Badler, N. I. (2009). How the ocean personality model affects the perception of crowds. *IEEE Computer Graphics and Applications*, *31*(3), 22–31.

González, I., Déjean, S., Martin, P. G., & Baccini, A. (2008). Cca: An r package to extend canonical correlation analysis. *Journal of Statistical Software*, *23*, 1–14.

Hernandez, H. (2021). Testing for normality: What is the best method. *ForsChem Research Reports*, *6*, 2021–05.

Mullola, S., Hakulinen, C., Presseau, J., Gimeno Ruiz de Porras, D., Jokela, M., Hintsa, T., & Elovainio, M. (2018). Personality traits and career choices among physicians in finland: employment sector, clinical patient contact, specialty and change of specialty. *BMC medical education*, *18*, 1–12.

5 Appendices

A	B	C	D	E	F	G	H	I	J	K
O_score	C_score	E_score	A_score	N_score	Numerical Aptitude	Spatial Aptitude	Perceptual Aptitude	Abstract Reasoning	Verbal Reasoning	Career
5.45	8.67	3.45	5.34	4.23	9.23	4.56	6.78	7.89	6.12	Accountant
8.78	5.67	4.56	6.45	4.23	5.12	8.45	7.89	6.34	6.01	Graphic Designer
6.12	6.78	9.34	7.56	5.01	6.23	4.23	6.45	6.67	8.45	Salesperson
9.12	8.78	4.23	5.67	4.56	7.89	5.34	6.45	9.34	7.67	Research Scientist
6.45	7.56	5.67	9.12	4.23	5.34	4.01	6.23	5.78	8.67	Teacher
8.45	7.01	5.34	5.45	4.56	6.45	9.12	7.67	8.89	4.23	Architect
4.01	8.23	6.67	9.34	5.12	4.56	4.45	7.45	5.78	7.78	Nurse
8.78	7.89	5.67	6.01	4.67	8.45	4.23	5.34	9.23	6.45	Software Developer
8.78	7.45	5.34	8.45	6.01	4.56	4.01	4.23	6.67	9.23	Psychologist
6.78	6.45	8.45	7.67	4.56	3.45	3.12	9.34	3.67	4.23	Chef
6.23	7.34	8.78	7.45	5.67	5.12	4.34	4.67	6.01	7.89	Marketing Manager
7.89	9.23	5.01	7.45	6.34	4.78	4.01	4.56	5.67	8.67	Physician
9.34	3.56	7.23	6.78	5.45	2.89	2.34	9.12	5.23	4.01	Artist

Figure 7: Part of the data set

R codes used for analysis:

```
library(tidyverse)
library(dplyr)
library(GGally)
library(CCA)
library(CCP)
data = read_csv("Data.csv") data % %summarise(across(everything(),    sum(is.na(.))))

DuplicateRows = data %  % filter(duplicated(.))
count(DuplicateRows) data_for_CCA = data[,-11] shapiroResults = sapply(data_for_CCA, shapiro.test)


setA = data_for_CCA[,1:5]

setB = data_for_CCA[,-(1:5)] ggpairs(setA)

ggpairs(setB) matcor(setA,setB) cc1 = cc(setA,setB)


cc1$cor
sqrt(cc1$cor)
cc1$xcoef
cc1$ycoef cc2 = comput(setA,setB, cc1)

cc2[3:6] rho = cc1$cor n = dim(setA)[1]


p = length(setA)
q = length(setB) p.asym(rho, n, p, q, tstat = "Wilks")

p.asym(rho, n, p, q, tstat = "Hotelling")
```