Brazilian Admission Exam for the Diplomatic Career: an Application of the Principal Components Analysis

Cauan Braga da Silva Cardoso*
University of Brasilia, Brasilia, Brazil – cauan.br@gmail.com.br

This study aims to analyze the Brazilian Diplomatic Career Admission Exam (CACD) of the Rio Branco Institute (IRBr) of five years: 2010, 2011, 2012, 2013 e 2014. The IRBr is the Brazil's diplomatic academy and organizer of the CACD, in partnership with the Center for Selection and Event Promotion of the University of Brasilia (CESPE/UnB). The CACD is one of the most grueling and selective exam of Brazil and consists of 10 writing exams divided in 4 phases until 2013 and in 3 phases in 2014. The approbation in CACD allows the candidate to ingress in the initial position of the diplomatic career (Third Secretary). The Principal Components Analysis (PCA) permits to demonstrate the relevance of each exam in the candidate final classification. In addition, the biplot of the first and second principal component provides a useful tool of data analysis and allows the visual appraisal of the influence of each exam between the candidates approved and reproved in the IRBr's entrance exam. The analysis of the CACD within the five consecutive years showed the importance of the English writing exam, followed by the Spanish and French writing exam, the Economy exam (in 2010 and 2011 only), and Brazil's History exam (in the following years).

Keywords: CACD; Principal Components; Biplot; R language.

1. Introduction

The Rio Branco Institute (IRBr) was founded in 1945 and is responsible for the selection and training of the Brazilians diplomats, in a continuous forming process, through the Formation Course, in the initial position of the diplomatic career (Third Secretary), the Improvements of Brazilian diplomats Course (CAD), for the Seconds Secretaries, and the High Studies Course (CAE) for Counselors.

The IRBr's entrance exam, or Admission Exam for the Diplomatic Career (CACD), take place annually since 1986 and counts with the participation of the Center for Selection and Event Promotion of the University of Brasilia (CESPE/UnB) as the organizer of the exams since 2002. The approval in the entrance exam allows the candidate to ingress in the initial position of the diplomatic career, according to the number of vacancies available, and to enroll in the IRBr's Formation Course.

Its current structure has 10 exams divided in 4 phases, in 2010, 2011, 2012 and 2013, and 3 phases in 2014:

- The first phase, also known as "Pre-Selection Exam" (TPS), is a qualifying test that involves the disciplines: Portuguese, Brazil's History, World's History, Geography, International Politics, English, Economy, Constitutional Law and Public International Law, with multiple choice questions and "True" or "False" questions.

- The second phase is an eliminatory and classificatory Portuguese writing exam, which consists of a dissertation about a general subject, with an extension from 600 to 650 words.

- The third phase is also an eliminatory and classificatory exam, composed by 6 exams applied in different days. They are, in order of application: Brazil's History, English, Geography, International Politics, Constitutional Law and Public International Law and Economy.

- The fourth phase is the only non-eliminatory phase and is composed by a Spanish and French exam.

The final ranking is given by the sum of the grades received in each exam between the candidates not disqualified in the eliminatory phases. The candidates occupy the vacancies according to the final classification; the vacancies reserved for the physically handicapped are occupied in the same way. *Table 1* gives the number of vacancies offered in each year.

Table 1 – Vacancies for regular candidates, physically handicapped candidates and total of vacancies for the CACDs from the years 2011 to 2014.

	Years				
Vacancy	**2010**	**2011**	**2012**	**2013**	**2014**
Regular	102	24	28	28	17
Quota for Physically Handicapped	6	2	2	2	1
TOTAL	108	26	30	30	18

The average number of candidates per vacancy in the five years considered was 203.86.

2. Methodology

For the analysis, the first step was to assemble the database with the grades in each of the 10 exams of the 4 phases (3 phases in 2014) of all candidates approved in the first phase (and not disqualified in the second or third phase), ordered by the final ranking, in five matrices, one for each year. The physically handicapped candidates were not considered for the analysis.

In the 2010 and 2011 CACDs, the TPS (first phase) selected the top 300 among the applicants, but only 165 and 173, respectively, made it to the fourth phase; in the 2012 CACD, the TPS selected the top 200 and 163 made it to the fourth phase; in the 2013 and 2014 CACDs, the TPS selected the top 100 and 91 and 100 made it to the fourth phase, respectively.

An analysis of principal components (ACP) often reveals relationships that were not previously suspected and thereby allows interpretations that would not ordinarily result (JOHNSON & WICHERN, 2007). The 10 exams formed the *p* random variables $X_1, X_2, ..., X_p$, p = 1, …, 9 (**Table 2**). In 2014, exceptionally, the Geography and International Politics exam were applied together and were considered as one variable for analysis purpose (**Table 3**). The principal components $Y_{1xp} = \{Y_1, Y_2, ..., Y_p\}$ are the uncorrelated linear combinations $Y_i = e'_{1xp} X_{px1}$ where $e'_{1xp} = \{e_1, e_2, ..., e_p\}$ is the eigenvectors of the covariance matrix $\sum$ of the random vector $X'_{1xp} = \{X_1. X_2 ...., X_p\}$, with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$. A principal component analysis is concerned with explaining the variance-covariance structure of a set of variables through a few linear combinations of these variables (JOHNSON & WICHERN, 2007).

Table 2 – Variables analyzed in the CACDs from 2010 to 2013.

X_i	VARIABLE	CODIFICATION
X_1	Pre-Selection Exam (TPS)	"tps"
X_2	Portuguese writing exam	"f2"
X_3	Brazil's History exam	"hb"

X_i		CODIFICATION
X_4	English writing exam	"ing"
X_5	Geography exam	"geo"
X_6	International Politics exam	"pi"
X_7	Constitutional Law and Public International Law exam	"dir"
X_8	Economy exam	"eco"
X_9	Spanish and French exam	"esp/fr"

Table 3 - Variables analyzed in the 2014 CACD.

X_i	VARIABLE	CODIFICATION
X_1	Pre-Selection Exam (TPS)	"tps"
X_2	Portuguese writing exam	"f2"
X_3	Brazil's History exam	"hb"
X_4	English writing exam	"ing"
X_5	Geography and International Politics exam	"geo/pi"
X_6	Constitutional Law and Public International Law exam	"dir"
X_7	Economy exam	"eco"
X_8	Spanish and French exam	"esp/fr"

The biplot provides a useful tool of data analysis and brings a visual elucidation of the structure of large data matrices. It is especially revealing in principal components analysis, where the biplot can show inter-distances and indicate clustering of units as well as displays variances and correlations of the variables (GABRIEL, 1971).

All the analysis was made using the R software, a free programming language and software environment for statistical computing and graphics.

3. Results

3.1 ACP of the 2010 CACD

The first principal component explain 29% of the total variance. The first two principal components, collectively, explain 51% of the total variance. The first three principal components explains 64% of the total variance. Consequently, the total variance is well summarized by the three principal components and a reduction in the data from 9 variables on 3 principal components is reasonable.

The first component appears to be essentially a weighted sum of the English writing exam and Spanish and French exam. The second component gives more weight to the Economy exam. The third component gives more weight to the Spanish and French exam.

3.1 ACP of the 2011 CACD

The first principal component explain 23% of the total variance. The first two principal components, collectively, explain 46% of the total variance. The first three principal components explains 61% of the total variance. Consequently, the total variance is well summarized by the three principal components and a reduction in the data from 9 variables on 3 principal components is reasonable.

The first component gives more weight to the Economy exam. The second component is a weighted sum of the English writing exam and Spanish and French exam. The third component gives more weight to the Geography exam.

3.1 ACP of the 2012 CACD

The first principal component explain 29% of the total variance. The first two principal components, collectively, explain 49% of the total variance. The first three principal components explains 65% of the total variance. Consequently, the total variance is well summarized by the three principal components and a reduction in the data from 9 variables on 3 principal components is reasonable.

The first component appears to be essentially a weighted sum of the English writing exam and Spanish and French exam. The second component is a weighted difference between the Brazil's History exam and the English writing exam. The third component gives more weight to the Spanish and French exam.

3.1 ACP of the 2013 CACD

The first principal component explain 29% of the total variance. The first two principal components, collectively, explain 51% of the total variance. The first three principal components explains 66% of the total variance. Consequently, the total variance is well summarized by the three principal components and a reduction in the data from 9 variables on 3 principal components is reasonable.

The first component appears to be essentially a weighted sum of the English writing exam and Spanish and French exam. The second component gives more weight to the English exam. The third component is a weighted difference between the Brazil's History exam and the Spanish and French exam, as show in the Graphic 1.

3.1 ACP of the 2014 CACD

The first principal component explain 45% of the total variance. The first two principal components, collectively, explain 65% of the total variance. The first three principal components explains 77% of the total variance. Consequently, the total variance is well summarized by the three principal components and a reduction in the data from 9 variables on 3 principal components is reasonable.

The first component gives more weight to the English writing exam. The second component is a weighted sum of Brazil's History exam and the Geography exam. The third component is a weighted difference between the Spanish and French exam and the Geography exam.

4. Conclusions

As the number of candidates and the number of vacancies were gradually decreasing over the years, more evident become the foreign languages exams. The biplot of the CACDs can show this growth in influence on the final ranking comparing the candidates approved and not approved in the IRBr's entrance exame (**Figure 1**).

On the following biplots, the numbers represent the candidates approved and its final classification. The "dots" are the not approved candidates. The horizontal axis is the first principal component and the vertical axis is the second principal component.

Figure 1 - Biplot of the CACD (2012-14).
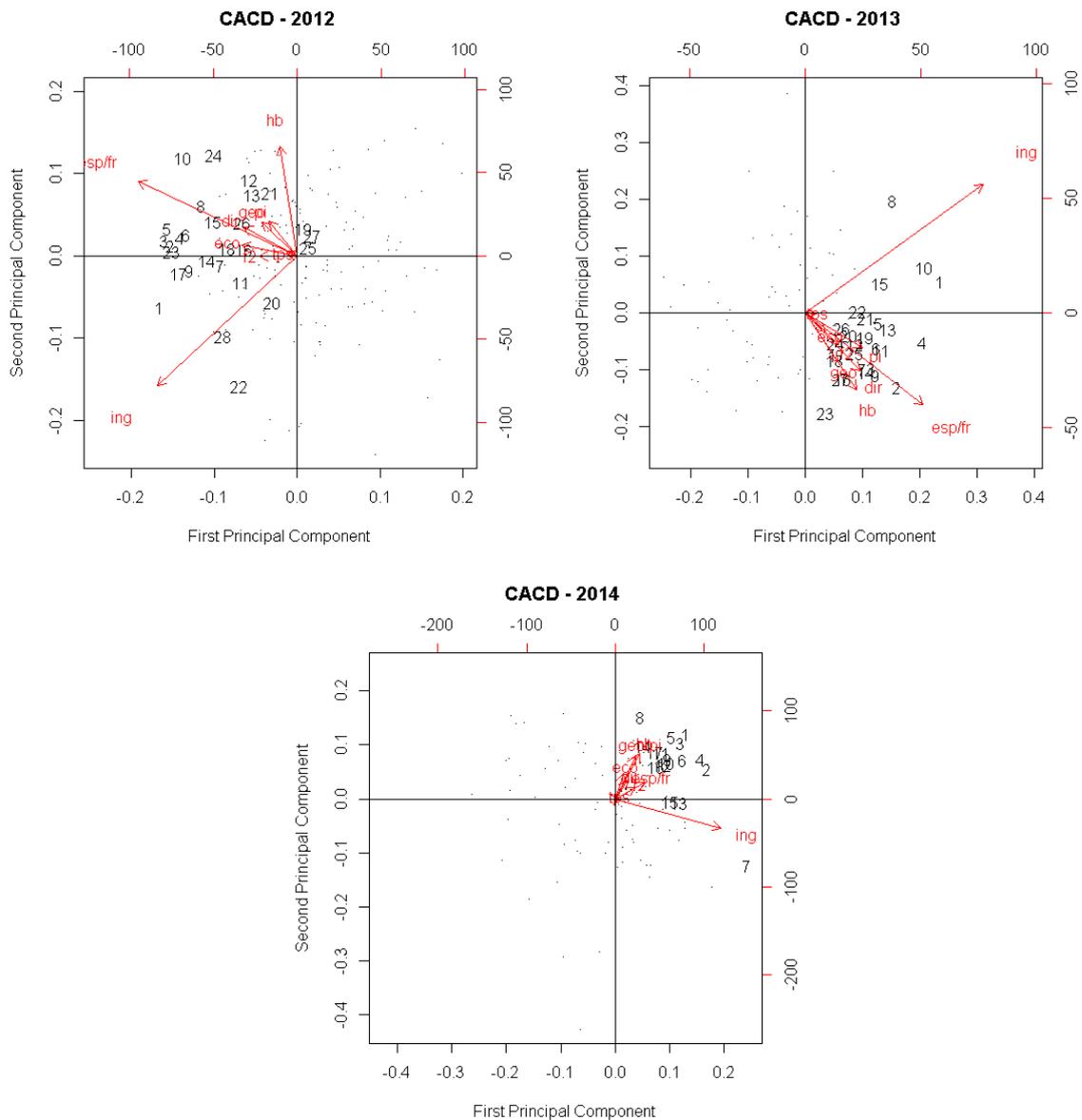
Besides the little difference between the final grades of the candidates, it still can be more evident through data reduction with PCA.

References

GABRIEL, K. R. (1971). The Biplot Graphic Display of Matrices with Application to Principal Component Analysis. *Biometrika*, **58**, pp. 453-467. Oxford: Biometrika Trust.

MINISTÉRIO DAS RELAÇÕES EXTERIORES. (2002). Instituto Rio Branco. [ONLINE] Available at: http://www.institutoriobranco.mre.gov.br/pt-br/Main.xml. [Accessed 22 February 15].

JOHNSON, R. A.WICHERN, D.W. (2007). Applied multivariate statistical analysis, 6th ed., New Jersey: Prentice Hall.