Principal Component Analysis (PCA) Technique for Finding the Best Applicant for a Job (Case study-Cihan University-Erbil)

Dr. Abood Mohammed Jameel, PhD
Cihan University-Erbil

Abstract

- **Abstract**
- This paper focuses on the use of principal component analysis technique (PCA) in choosing the best applicant for a job in Cihan University-Erbil. Cihan University has a panel of judges (University staff) to help in choosing the applicants for a job by evaluating or rating each one on different scale of preference and different type of characteristics. This process usually creates complicated multivariate data structure, which consists of 25 applicants for a

Cont.

for a job rated by a panel of judges on 17 characteristics [25 rows, applicants, and 17 columns, characteristics]. PCA plays a crucial role in conducting impactful research as it offers a potent technique for analyzing multivariate data. Researchers can utilize this method to extract valuable information that aids decision-makers in problem-solving. To ensure the appropriateness of data for PCA, certain testing procedures are necessary. In this study, two tests, namely the Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy and Bartlett's Test of Sphericity, were performed, and their significance is vital. The findings indicate that the data employed in this research are suitable for PCA. Scoring and ranking procedures as extra tools were used to see that applicant No. (1) is the first accepted for a job, applicant No. (17) is the second, applicant No. (12) is the third, and so on.

1.1 Introduction

Cihan University is a new private university in Erbil-Kurdistan. Cihan University was established in 2007 to help students to get B.Sc. degree in different fields. This university was started with three departments but now more than 25 departments. The university has a panel of judges (University staff) to help in choosing the applicants for a job by evaluating or rating each one on different scale of preference and different type of characteristics.

3. Research Importance

The procedure of selecting or choosing the best applicant for a job depends most of times on subjective opinions. Using PCA technique gives the judges scientific procedure to help them in choosing the best applicant for a job.

4. Research Objectives

- The main objective of this paper is to find the most preferable applicant for a job in Cihan University-Erbil by using PCA, allocating scores and ranks.

Literature Review

PCA is a data processing technique used to extract a limited set of composite variables, known as principal components, from a larger set of measured variables. These principal components aim to capture and explain a specific phenomenon (Hastie et al., 2009; Constantin, 2014).

PCA is widely acknowledged as a valuable technique for reducing dimensions and compressing data. It generates orthogonal factors that account for a significant portion of the variation observed in the variables meeting specific criteria (Hastie et al., 2009).

Cont.

Various extraction rules and methods exist for determining the number of factors to retain. Among these, Kaiser's criteria are widely recognized and suggests retaining only those factors with eigenvalues greater than 1. Another approach involves using the Scree plot or the cumulative percentage of extracted variance (Williams et al., 2010).

Application

3.1 Data Sources

We gathered 25 forms from a panel of judges who interviewed 25 job applicants at Cihan University-Erbil. The judges rated the applicants on various scales for each of the 17 described characteristics listed on the form (as in Table 1).

Table 1
The variable's description

V.	Description	V.	Description
X_1	General Appearance, Health, and Build, 1-3	X_{10}	Other Languages, 0-5
X_2	Dress; Formal, Informal, and Clean, 1-3	X_{11}	Personality: Weak, Hesitate, Strong, 2-6
X_3	Voice: Too loud, Too Low, Pitchy, 1-3	X_{12}	Leadership: Ability to control, 0-12
X_4	Lecturing Experience: one mark for each year, max. 6.	X_{13}	Communication, 0-10
X_5	Job experience: one mark for each year, max. 7	X_{14}	Presentation: stability, lecture method, lecture subject, 0-10
X_6	Academic Rank: A. Lect. , Lect., A. Prof., Prof., 6, 8, 10, 12	X_{15}	Time punctuality: On time: 2, $\geq$ 15 mint: 0, $<$15 mint: 1
X_7	Researches: One mark for each research, max. 5	X_{16}	Arabic Languages: 0-10
X_8	English Languages, 0-10	X_{17}	Department Recommendation: 0-4

Table 1
The variable's description

The measurements are organized in a table or matrix format with 25 rows and 17 columns, as illustrated in Table 2.

Table 2

Interview Assessment from Judgment Panel of Judges - Cihan University-Erbil

No.	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	Total
1	2	3	4	6	5	8	2	8	5	2	4	4	8	7	2	10	3	83
2	2	3	3	5	4	6	2	6	5	2	4	4	8	6	2	10	2	74
3	2	3	4	6	6	8	2	8	5	2	4	5	7	8	2	8	2	82
4	3	2	4	5	5	6	3	8	5	3	5	4	8	8	2	10	3	84
5	2	2	4	4	5	6	2	6	6	2	4	5	7	7	2	8	3	75
6	3	3	3	5	5	6	3	6	5	2	5	5	8	7	2	8	3	79
7	2	2	3	4	5	6	2	6	5	2	4	4	6	6	2	8	2	69
8	3	2	2	4	4	6	2	6	5	2	4	4	6	6	2	8	3	69
9	3	3	4	4	5	8	2	6	5	3	5	5	8	6	2	10	3	82
10	2	2	3	3	4	6	2	6	6	2	4	4	6	7	2	8	2	69
11	2	3	3	4	3	6	3	6	4	3	4	4	6	6	2	8	3	70
12	3	3	4	5	5	8	3	6	5	3	5	5	7	8	2	10	3	85
13	3	2	3	4	4	6	2	6	5	2	4	4	6	7	1	8	2	69
14	3	3	4	4	4	6	3	6	5	3	4	5	6	6	1	10	2	75
15	2	2	3	4	4	7	2	6	4	2	4	4	6	6	2	8	2	68
16	3	2	2	4	4	6	2	6	5	3	4	5	6	6	1	8	2	69
17	3	3	4	5	6	7	3	6	5	3	5	5	8	7	2	10	3	79
18	2	3	4	4	3	6	2	4	4	2	4	4	6	7	2	8	3	68
19	3	2	4	5	4	5	3	4	5	3	4	4	5	6	2	8	2	69
20	3	2	4	4	4	6	3	5	5	3	4	5	5	6	2	10	3	78
21	2	2	3	4	5	5	3	5	6	3	5	4	5	6	2	8	2	70
22	3	3	4	5	6	6	3	6	5	3	4	5	7	7	2	10	3	82
23	2	3	4	4	5	6	3	6	5	3	5	5	7	8	2	10	3	81
24	3	3	4	5	6	5	3	6	6	3	4	5	8	8	2	8	3	82
25	2	3	3	4	6	6	3	5	5	3	4	4	7	7	2	8	2	74

3.2 Data Analysis

In order to examine a data set, similar to the one found in table 2, we employ the PCA technique. Before commencing our data analysis, it is crucial to perform two testing procedures to ascertain whether the data is appropriate for this method. To accomplish this, we utilize the following two tests.

1. Kaiser-Meyer-Olkin (KMO) index or Measure of Sampling Adequacy (Williams et al. 2010). The KMO index is a numerical value that falls between 0 and 1. If this index is equal to or greater than 0.50, it indicates that the sample is appropriate for performing PCA

2. Bartlett's Test of Sphericity to be considered significant ($p < 0.05$) according to Constantin (2014), the analysis is conducted using the JAMOVI 2.2.5 package from the Jamovi project (2021). The obtained results are as follows:

KMO = 0.561, it is greater than 0.50, hence the sample is considered suitable for PCA.

Furthermore, the Bartlett's test of Sphericity shows a significance level of ($p<0.002$) with degrees of freedom (df) equal to 136, indicating a highly significant result. The findings are presented in Table 3 indicate that the data utilized in our example are suitable for conducting PCA.

Table 3
KMO and Bartlett's Test

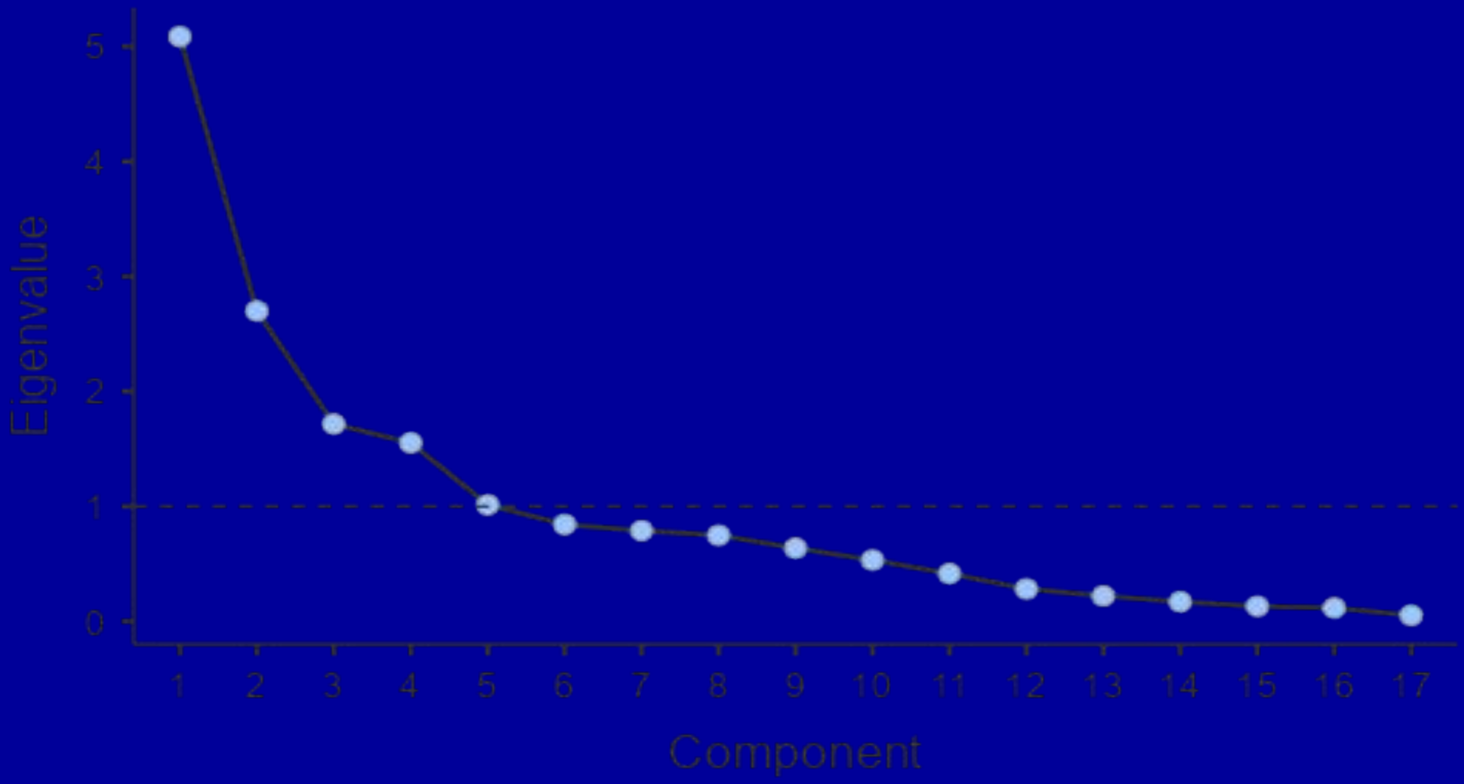
The findings presented in Table 3 indicate that the data utilized in our example are suitable for conducting PCA.

Table 3
KMO and Bartlett's Test

Bartlett's Test of Sphericity			KMO Measure of Sampling Adequacy (MSA)
χ^2	df	p	Overall 0.561
188	136	< .002	

KMO and Bartlett's Test

Fig. 1
Illustrating the scree plot for the original variables



- Following these two tests, we can employ PCA to determine the appropriate number of principal components to retain in the model. Initially, the number of components matches the number of variables (17 variables) included in the model. Each component possesses an eigenvalue, which indicates the amount of variance explained by that particular component. The eigenvalues and eigenvectors of the correlation matrix were derived using JAMOVI package.

•The first principal component accounted for 28.077% of the total variance; the second a further 19.413 %; the third a further 10.004 %; the fourth 9.153 %; the fifth 6.056% making 72.7 % of the total variance "explained" by five uncorrelated combinations of the original variables (see Table 4).

- The Kaiser's criterion, also referred to as the eigenvalue-one criterion, is a widely used approach for selecting principal components. According to this criterion, only variables with eigenvalues greater than 1 are retained in the new model. As a result, the 6th variable, which has an eigenvalue of 0.8200 (refer to Table 4), will be excluded from the model.

Table 4
The amount of variance accounted for (Initial Eigenvalues)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.7731	28.077	28.1	4.7731	28.077	28.1
2	3.3002	19.413	47.5	3.3002	19.413	47.5
3	1.7007	10.004	57.5	1.7007	10.004	57.5
4	1.5560	9.153	66.6	1.5560	9.153	66.6
5	1.0295	6.056	72.7	1.0295	6.056	72.7
6	0.8200	4.823	77.5			
7	0.7747	4.557	82.1			
8	0.6435	3.785	85.9			
9	0.5833	3.431	89.3			
10 11 12 13	0.5284 0.3625 0.2870 0.2214 0.1850 0.1378	3.108 2.132 1.688 1.302	92.4 94.5 96.2 97.5			

Table 5
Eigenvectors for the first five components (Component Loadings)
Note: 'varimax' rotation was used

Variable	Component				
	1	2	3	4	5
X1	0	0	0	0	0
X2	0.827	0	0	0	0
X3	0.668	0	0.302	0	0
X4	0.727	0	0	0	0
X5	0.446	0	0	0	0.751
X6	0.395	0.699	0	0	0
X7	0.332	-0.728	0	0	0
X8	0	0.741	0	0	0
X9	0	0	0	0	0.855
X10	0	-0.623	0	0.529	0
X11	0	0	0.763	0	0
X12	0.343	0	0	0.609	0
X13	0.540	0.454	0.371	0	0.315
X14	0.499	0	0	0	0.499
X15	0	0	0.550	-0.678	0
X16	0.431	0	0.470	0.443	0
X17	0	0	0.774	0	0

- **Now we are not interested only in the interpretation of the components in this research; but we wish also to consider the component of scores which can be produced by post-multiplying the original data matrix (25×17) by the matrix of eigenvectors (17×5) using (Mathcad 15 M050) software, due to the large size of the matrices. This process produce and allocate score to each applicant, and then researchers find the component of ranks by ranking the score component as described in table 6 below.**
- **The result of the previous procedure analysis reveals that the applicant No.1 has the greatest score (29.064), and ranked No.1 and the applicant No. 17 has score (29.063) and ranked No.2, and so on (see Table 6).**

Result of Statistical analysis

Applicant No.	Score	Rank	Applicant No	Score	Rank	Applicant No	Score	Rank
1	29.064	1	6	26.692	10	13	22.937	19
17	29.063	2	2	25.934	11	7	22.884	20
12	28.971	3	25	25.528	12	15	22.833	21
3	28.95	4	14	25.47	13	21	22.281	22
22	28.128	5	5	24.934	14	10	22.21	23
24	27.91	6	20	24.103	15	16	22.113	24
4	27.551	7	18	23.986	16	8	21.77	25
23	27.454	8	19	23.23	17			
9	27.454	9	11	23.151	18			

Conclusions

. The process of allocating scores and ranks of the previous procedure analysis reveals that the applicant No.1 has the greatest score (29.064) ranked No.1, the applicant No. 17 has score (29.063) ranked No.2 and applicant No.12 has score (28.971) ranked No.3. And so on (see Table 5). Finally, conclude that this PCA is a very good and power full scientific technique can beused to help in analyzing complicated data structure to give advice in selecting the best applicant for a job.