Al Kut Journal of Economics and Administrative Sciences /ISSN: 1999 -558X /ISSN Online 2707-4560/ Vol (18) Issue: 61-2026 (March)

AL KUT JOURNAL OF ECONOMIC AND ADMINISTRATIVE SCIENCES
Publisher: College of Economics and Management - Wasit University

Using Factor Analysis to Extract The Most Important Factors Influencing The Economic Activity of large Industrial Establishments in Iraq in 2020

Mustafa Habib Mahdi

Middle Technical University , Institute of Administration/ Al Rusaffa , Department of Statistics Techniques. Baghdad, Iraq.

mustafa_habib@mtu.

Abstract

Principal Component Analysis (PCA) is one of the most important statistical methods in factor analysis. It is used to reduce data dimensions and extract the underlying factors that represent the underlying structure of variables. This research aims to study the effectiveness of the principal components approach in anal yzing data from a sample of (719) large industrial establishments in Iraq in 2020, using the JASP program .The study relied on a set of basic steps applied to principal components analysis, which included (data preparation and normalization, calculating the variance-covariance matrix, extracting the principal components using singular value decomposition (SVD), and interpreting the extracted factors in light of the explained variance).

The results showed that the principal components analysis method clearly contributed to simplifying the data structure by reducing the number of input variables while retaining most of the total variance, allowing for a clearer and more accurate interpretation of the factors influencing the performance of large industrial establishments. This approach also helped uncover common patterns among variables and obtain a more comprehensive picture of the economic data under study.

Keywords: Factor analysis, principal components method, large industrial establishments in Iraq in 2020.

1. Introduction

The industrial sector is one of the most important productive sectors in any country's economy, given its prominent role in establishing the material foundation for progress and its ability to determine the desired path of growth in various economic, political, and social fields. From this perspective, developing the industrial sector represents a strategic goal for achieving desired economic, political, and social objectives, as it is a fundamental axis for driving development.

Advancing industry also contributes to achieving high economic growth rates, providing widespread job opportunities, and producing goods locally rather than relying on imported goods. This leads to increased economic diversification, which is essential for promoting social and economic growth in various countries. Given this importance, the Central Agency for Public Mobilization and Statistics works to provide comprehensive statistical data on the reality of industrial activity and provides accurate information for decision-makers to formulate appropriate policies to raise the industrial sector's input to the overall domestic product [5].

This study focused on the attributes of the labor power in economic and social development, to determine the size and type of future economic activity in Muthanna Governorate. The study demonstrated a strong relationship between some demographic variables (such as fertility, household income, marital status, and education) and the labor force, as these variables explained approximately (94.9 %) of the phenomenon studied [2].

This study aims to determine the impact of Hamma Bouziane Cement Company's commitment to environmental responsibility on its image among stakeholders, including residents and employees. The study reached several conclusions, the most important of which is that Hamma Bouziane Cement Company provides its employees with important and necessary information about its environmental commitment, which helps it build a positive image of itself in its internal and external environment [8].

This study Checked the most important factors influencing employee function satisfaction, using exploratory factor analysis using the principal components method. It concluded that four factors influence employee job satisfaction: interactions with colleagues, job satisfaction, managerial interactions, and the work environment.[1]

The study evaluated the SVDU-IPCA method using general databases, any FERET, AR, and Yale B, and applied it for algorithms. Experimental outcomes showed that the variation in average identification precision between this method and the principal component analysis (PCA) method is less than 1%.[3]

2. Research Methodology

2 1. Research Problem.

Due to the large number of economic variables and their intertwining with each other, which made the analysis complex, the results will therefore be unstable .

2.2. Research Objective.

Using the Principal Component Analysis (PCA) method to extract a limited number of independent components that represent the basic dimensions of economic activity such as (production, sales, wages of permanent and temporary workers, ..., etc.) instead of dealing with each (18) variable separately .

3. Factor Analysis

The factor model for k observed variables for a sample of size n is explained based on a linear function of m common factors, where k > m individual factors for each variable, i.e.,:

$$\underline{Y} = B\underline{F} + \underline{U} \qquad (1)$$

Where:

k*1: vector of variables of degree (Y)

k*m: matrix of factor loadings of degree (B)

m*1: vector of common factors of degree (F)

k*1: vector of individual factors of degree (U).

The principal component (PC) method is one of the main methods in factor analysis. It explains a phenomenon that relies on a large number of independent variables to achieve the highest degree of information, with fewer independent factors than the variables used, which express the existing relationships between the variables. [4][7]

To find the matrix T, $T'ST$, where D is a diagonal matrix whose diagonal elements represent the characteristic roots of the matrix S. T is an orthogonal matrix whose columns represent the orthogonal characteristic vectors. The first principal component V_1 of the original variables is a linear combination, where:

$$V_1 = b_{11}Y_1 + b_{21}Y_2 + \cdots + b_{P1}Y_P \qquad (2)$$

$$= \underline{b}_1' \, Y_\infty$$

And b_1 is the characteristic vector corresponding to the characteristic root λ_i, assuming that:

$$V_1' \approx Y_1 N_P(0, \underline{b}_1' S_{\underline{b}_1})$$

Since $S_{\underline{b}_1} = \lambda_i$, $V_1 = N_P(0, \lambda_1)$, that is, the largest characteristic root of the matrix is used to estimate the maximum variance of the first principal component, and the characteristic vector b_1 is used to estimate the first principal coefficients. Similarly, the second characteristic root of the matrix S is used to estimate the maximum variance of the second principal component. In general, $\underline{V} = \Gamma \underline{Y}$, and the covariance between V_2, V_1 is COV(V_2, V_1), that is, the correlation between the first principal component and the second principal component is zero. Each factor analysis method has several considerations in determining the basis for determining the number of significant factors, including the method developed by Kaiser in 1950.

$$Y\underline{b_i} = \lambda_i \underline{b_i} \qquad (3)$$

$$(Y - \lambda_i I)b_i = 0 \qquad (4)$$

$$|Y - \lambda_i I| = 0 \qquad (5)$$

$$Y\underline{b_i} = \lambda_i \underline{b_i} \qquad (6)$$

$$S(b_{ip}) = \left(S(r_{xy})\right)\sqrt{{}^{k}/_{k+1-p}} \qquad (7)$$

$$KMO = \frac{\sum_{i \neq j} \sum r_{ij}^2}{\sum_{i \neq j} \sum r_{ij}^2 + \sum_{i \neq j} \sum b_{ij}^2} \qquad (8)$$

4. Singular Value Decomposition (SVD)

is the process of decomposing a matrix (either square or rectangular) into the product of three specific matrices :[6]

$$A = U\, \Sigma\, V^T \qquad (9)$$

Where:

A: is the original (m×n) data matrix to be decomposed.

U: is an orthogonal (m×m) matrix called the left singular vector.

Σ: is an (m×n) diagonal matrix called the singular values matrix, containing the singular values (non-negative real numbers in descending order) on its main diagonal.

V^T : It is the transpose of an orthogonal matrix (n×n), whose columns are called right singular vectors.

The singular values in the Σ matrix are a measure of the importance of each singular vector in describing the variance within the data.

Singular value decomposition (SVD) is a basic and effective computational method for extracting principal components from a data matrix.

5. Results

From the results of Table (1) and (2), it is clear that the sample size is suitable for using agent analysis (principal components method (PCA)), as the value reached (KMO(MSA) = 0.684) which is greater than the minimum (0.500), and the value of (Bartlett’s test) reached (X2 = 17132.601) with a probability (P) less than (0.001), which indicates that the correlations between the variables are significant, meaning that it is possible to extract the factors.

Table (1) shows the suitability of the sample size for analysis using the Kaiser-Meyer-Olkin test	
	MSA
Overall MSA	**0.684**
Number of months of work	0.518
Number of permanent employees	0.648
Wages of permanent employees	0.776
Average salary	0.901
Number of temporary workers	0.346
Temporary workers' wages	0.744
Average wages for temporary workers	0.468
Total number of employed people	0.649
Unpaid workers	0.598
Production value at market price Product price	0.780
Non-commodity production value and revenues	0.606
Sales value	0.727
Net sales of goods for resale	0.514
Raw materials and primary materials	0.790
Packaging materials	0.361
Other commodity requirements	0.973

Table (1) shows the suitability of the sample size for analysis using the Kaiser-Meyer-Olkin test	
	MSA
Service requirements	0.687
cost of goods for sale	0.513

Table (2) shows the significance of correlations by chi-square test			
	Value	**Df**	**p**
Model	17132.601	73	< 0.001

From Table (3), we conclude that the number of extracted components (eigenvalues) with values greater than one is (5). Together, these five components explain approximately (76.3%) of the total variance after the Varimax rotation. This is excellent because it reduces a large number of variables to only (5) factors.

After rotation, the factors emerged as follows:

- The first agent included the highest loadings of the variables (production value at market price = 0.941, sales value = 0.935, raw materials and primary materials = 0.888, other commodity requirements = 0.877, packaging materials = 0.627), and can be called the (production or commodity laborer).

  (Production, Sales, Raw Materials)

- The second laborer included the highest variable loadings (permanent employee wages = 0.737, value of production and non-commodity revenues = 0.531, number of permanent employees = 0.698, total number of employees = 0.693, average wage = 0.586), and could be called the "permanent employee wages worker ".

- The third factor included the highest variable loadings (service requirements = 0.694, net merchandise sales = 0.978, cost of goods sold = 0.978), and could be called the "trade or resale activity worker ".

The fourth factor included the highest variable loadings (temporary employee wages = 0.864, number of temporary employees = 0.847), and could be called the "temporary employee wages factor." The fifth factor contained the highest levels of variable influence (number of months worked = 0.804, unpaid workers = -0.795), and could be called the "unpaid labor factor." In contrast, one variable emerged as a standalone variable (mean wages of temporary workers = 0.857), as it was not strongly related to the other factors.

Table (3) shows the loadings of the five components

	PC1	PC2	PC3	PC4	PC5	Uniqueness
Production value at market price Product price	0.941					0.030
Sales value	0.935					0.029
Raw materials and primary materials	0.888					0.184
Other commodity requirements	0.877					0.150
Packaging materials	0.627	- 0.541				0.259
Service requirements	0.553		0.694			0.172
Wages of permanent employees	0.547	0.737				0.095
Non-commodity production value and revenues	0.522	0.531				0.403
Number of permanent employees	0.444	0.698				0.156
Total number of employed people	0.429	0.693		0.422		0.133

Table (3) shows the loadings of the five components						
	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**Uniqueness**
Average salary		0.586				0.626
Net sales of goods for resale			0.978			0.037
cost of goods for sale			0.978			0.037
Temporary workers' wages				0.864		0.161
Number of temporary workers				0.847		0.233
Number of months of work					0.804	0.349
Unpaid workers					-0.795	0.359
Average wages for temporary workers						0.857
Note. Applied rotation method is varimax.						

Table (4) shows the variances of the five factors before rotation. It was found that the first factor explained (36.7%) of the variance, knowing that (eigenvalue = 6.607), the second factor (49.87%), i.e. by adding (13.1%), the third factor (62.2%), the fourth factor (69.4%), and finally the fifth factor (76.3%). As for the variances after rotation (varimax), they were as follows: The first factor decreased from (36.7%) to (27.8%), which indicates that rotation reduces the dominance of the first factor, while the results of the remaining factors were, in order: the second factor (15.9%), the third (13.6%), the fourth (11.6%), and the fifth (7.5%).

Table (4) shows the properties of the five components before and after recycling						
	Unrotated solution			**Rotated solution**		
	Eigenvalue	**Proportion var.**	**Cumulative**	**SumSq. Loadings**	**Proportion var.**	**Cumulative**
Component 1	6.607	0.367	0.367	4.997	0.278	0.278
Component 2	2.365	0.131	0.498	2.861	0.159	0.437
Component 3	2.228	0.124	0.622	2.442	0.136	0.572
Component 4	1.298	0.072	0.694	2.086	0.116	0.688
Component 5	1.231	0.068	0.763	1.343	0.075	0.763

The results in Table (5) are attributed to the use of orthogonal varimax rotation. This rotation aims to make the components statistically independent of each other, so that there is no significant overlap in interpretation. As a result, each component (RC1…RC5) explains an independent dimension of economic activity, and there is no linear relationship between the factors.

Table (5) shows the relationships between the components					
	Component 1	Component 2	Component 3	Component 4	Component 5
Component 1	1.000	0.000	0.000	0.000	0.000
Component 2	0.000	1.000	0.000	0.000	0.000
Component	0.000	0.000	1.000	0.000	0.000

3					
Component 4	0.000	0.000	0.000	1.000	0.000
Component 5	0.000	0.000	0.000	0.000	1.000

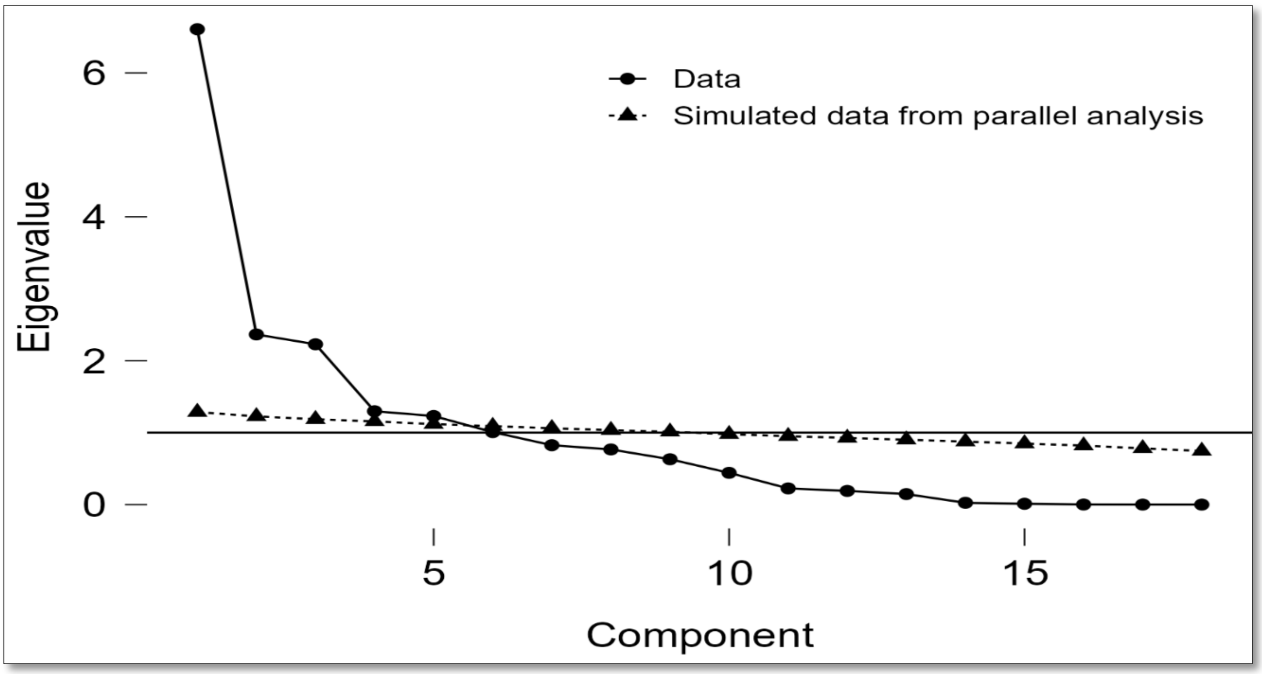

Figure (1) shows the Kaiser test and parallel analysis, which shows that there are approximately only (5) factors whose eigenvalues are higher than the values predicted from the parallel analysis. This is also consistent with Figure 2, which contains 5 factors (RC1 to RC5). After the fifth factor, the values begin to fall below the random line, indicating that the remaining factors do not provide a significant explanation for the variance.

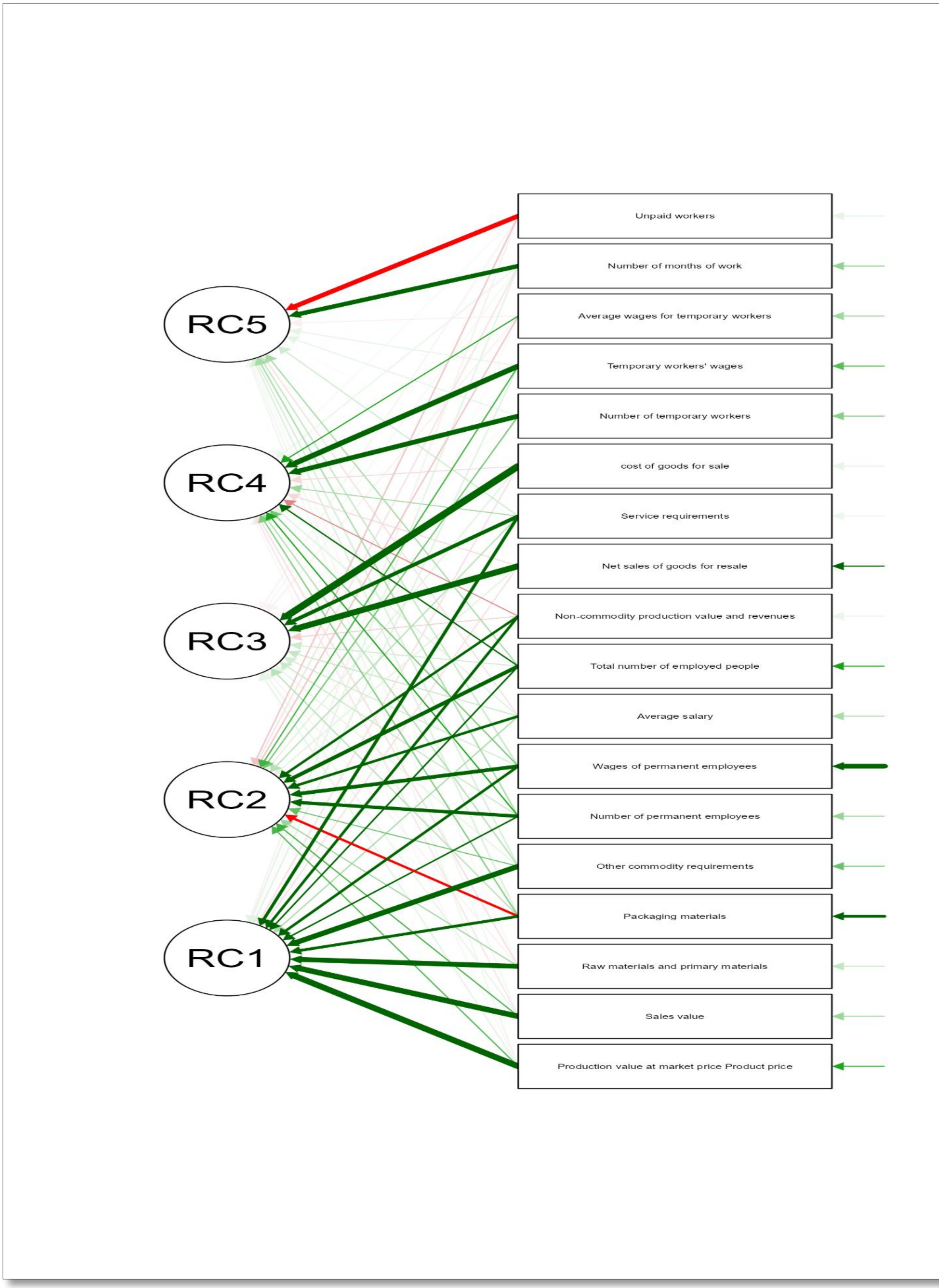
RC5
RC4
RC3
RC2
RC1
Unpaid workers
Number of months of work
Average wages for temporary workers
Temporary workers' wages
Number of temporary workers
cost of goods for sale
Service requirements
Net sales of goods for resale
Non-commodity production value and revenues
Total number of employed people
Average salary
Wages of permanent employees
Number of permanent employees
Other commodity requirements
Packaging materials
Raw materials and primary materials
Sales value
Production value at market price Product price

Figure (2) shows that the first factor (RC1) is closely related to the variables (production, wages, and number of workers). The second factor (RC2) is closely related to the variables (production requirements and raw materials). The third factor (RC3) is related to the variables of capital and loans. The fourth factor (RC4) is related to the variables of production costs and net revenues. The fifth factor (RC5) has a limited relationship with only one or two variables, and may be a secondary factor or specific to non-shared variables.

6. Conclusion

1. Number of primary factors (components): Five (5) were identified, focusing on high variability, with eigenvalues (eigenvalues) for each of which were greater than the true value (which is intended to retain the factor).
2. Proportion of total explained factors: These five factors together explained a significant proportion of the total variance in the aggregate industrial data, a total of 76.3%.
3. Effect of rotation on total variance: The total explained variance was not accumulated after rotation, with the percentage cut off at 76.3%. This confirms that rotation aims to redistribute the variance among the factors, not to change the total explained variance.
4. Interpretation of the principal factor (post-rotation): After analysis, the original components were grouped into five principal components, each with a clear explanation for their contribution and different selections:

   Factor 1: Production and Variety.

   Factor 2: New Workers and Wages.

   Factor 3: Resale or Trade.

   Factor 4: Temporary Workers and Wages.

   Factor 5: Working Time and Unpaid Workers.

7. References

[1] Abdul Wahab, Yasser Ismail Muhammad, Prof. Ahmed Hamad Al-Nouri (2020) “Using Exploratory Factor Analysis to Estimate and Analyze the Factors Affecting Job Satisfaction in Government Institutions,” Journal of Humanities and Natural Sciences.

[2] Al-Hasnawi, Jawad Kazim Obaid, and Al-Jayashi, Hamid and Kaa Sisan (2014). "Spatial analysis of the characteristics of the labor force and its impact on industrial development in Al-

Muthanna Governorate for the year 2014". Journal of Geographical Research, Issue (22), p. 217.

[3] Haitao Zhao, Pong Chi Yuen,James T. Kwok,(2006)."A Novel Incremental Principal Component Analysis and Its Application for Face Recognition",IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS-PART B: CYBERNETICS, VOL. 36, NO. 4.

[4] Issa, Aseel Muslim, Al-Rawi, Asmaa Ghaleb (2019) "A comparison between two methods of analyzing the main core components to reduce the image dimensions", Iraqi Journal of Statistical Sciences, Issue 29, pp. 11-24.

[5] Large Industrial Establishments Statistics Report (Cumulative), Ministry of Planning, Central Statistical Agency for the year 2020.

[6] Michael E. Wall , Andreas Rechtsteiner , Luis M. Rocha (2003). “Singular Value Decomposition and Principal Component Analysis”, Los Alamos National Laboratory, Cornell University, ed. 4.

[7] Muhammad, Haider Yahya, Al-Alawi, Liqa Ali Muhammad (2023). “Adapting Principal Component Analysis and Cluster Analysis with Practical Application”, Iraqi Journal of Economic Sciences, College of Administration and Economics, Issue Six, p. 1236.

[8] Yakhlef, Maryam Azza. (2021). "The Impact of Environmental Responsibility on the Image of Algerian Economic Institutions": A Field Study at the Cement Company of Hana Bouzian - SCHB Constantine. University of Debag Hine Setif 2.