Use factor analysis to study the factors affecting traffic accidents For the year 2018 in Iraq

L. Wafaa Abdul Samad Ashour

L. Mona Taher Ghafel

Department of Statistics - College of Management and Economics Department of Statistics - College of Administration and Economics

Summary

This research includes the application of one of the multivariate statistical methods is the factor analysis method that analyzes a set of correlation coefficients between several variables and reduces the variables under discussion to a smaller number resulting in new variables called factors' that is, it helps to understand the structure of the correlation matrix or common variation through Fewer factors.

Data were obtained from the annual statistical report issued by the Central Statistical Authority for the year (2019) and the research included (15) governorates and (16) variables representing the type of road, the time of the accident, and the type of accident. The research came out with a set of results, the most important of which are reducing the variables to three basic factors by Variation (85.82) of total variance.

Keywords: Global Analysis, ' Components Method,' Contributions of Variables.

The first topic: research methodology

1 -Introduction

The technological progress that prevails in all material, economic and cultural fields has provided people with many reasons for well-being. However, many of them did not make good use of these blessings, so they employed them negatively and threatened their lives and security. Traffic accidents are at the forefront of these negatives as a result of the resulting material and social damages and the harvest of the lives of many people. The danger of traffic accidents lies in the many effects that it has on the individual and society, including economic, social and psychological.

2 -Research problem

The increase in the number of traffic accidents in Iraq "and the resulting negative impacts are among the most important problems that are added to other problems, such as diseases and

drugs that kill human lives, which must be studied and try to put plans to reduce them" because of their contact and impact on human lives.

3 -Research objective

This research aims to use the factor analysis to study the most important factors affecting traffic accidents in Iraq for the year of 2018 using the statistical analysis package spss based on the most important elements that cause accidents, which are road type, lighting time and type of accident.

4 -The importance of research

The importance of the research emerges in adopting a statistical method (global analysis) to contribute to knowing the most important factors and causes leading to traffic accidents' in addition to raising interest to use various statistical methods to know the causes of traffic accidents to help specialists in developing future solutions to reduce accidents because of this issue's effects on development in Iraq.

5 -Previous studies

1- Bahjat Study (2010). "Using Factorial Analysis to Show Factors Affecting Cheating Among Students" This study concluded that there are six main factors for the phenomenon of cheating.

2 -Study (Nihad, 2012). "Using the global analysis to study the most important factors affecting child health in Salah al-Din Governorate / Iraq," and the study found that there are two main factors affecting child health.

3 -Study (Faraj and Khalil, 2014) "Using global analysis to determine the most important factors affecting the migration of Sudanese scientific competencies: a statistical economic study of the state of Sudanese scientific competencies in the Kingdom of Saudi Arabia", this study found that there are three factors that affect the brain drain.

4- Study (Kazem, 2014) "The role of factor analysis in determining the most important factors affecting the quality of health services provided to patients (Al Furat Central Teaching Hospital as a model)" in which two factors affecting the quality of health services were found.

5 -Study (Saad, 2016) "Factor analysis of the causes of academic failure for Dammam University female students". The study found that there are four factors behind the academic failure of university students.

6 -Study (Hussam, 2017) "Using factor analysis to determine the most important factors affecting the migration of Iraqi youth: a statistical study of the situation of youth migration in Al-Qadisiyah governorate." The study found that there are three basic factors affecting migration.

The second topic: theoretical side

Factor Analysis

It is a statistical method that works to shrink the number of variables or data related to a specific phenomenon 'This is done by building a set of specific variables on relationships and transforming them into a set of basic components that are not highly related among them. Where the best reconciliation of the basic components is the first factor' and also identifies reconciliation It is better for the second basic components that are not calculated in the first factor to determine the second factor and so on for the rest of the factors (Namik, p. 99.(

Types of factor analysis A distinction can be made between two types of factor analysis:

Exploratory Factor Analysis

This type is used in cases where the relationships between variables and the underlying factors are unknown and therefore the factor analysis aims to discover the factors to which the variables are classified.

Confirmatory Factor Analysis

This type is used for testing hypotheses regarding the presence or absence of a relationship between the variables and the underlying factors, as it is used as an elementary and empirical analysis in addition to evaluating the ability of the factors model to express the actual data set as well as in comparison between several models of factors in this field. [Salman ’p. 1]

General concepts in factor analysis

1- passive root Eigen Value

The sum of squares of the saturation of all variables on each factor of the array separately, and represents the amount of variance that the factor contributes to, and is determined by the value (1) so that if the latent root is greater than one we accept the factor and if it is less then we reject it.

2- Communality

The sum of the variable's contribution to the different factors that could be extracted in the global matrix 'and since one variable contributes different amounts to each factor' and whether its contributions are substantial or not significant 'then the sum of the squares of these contributions or the saturation of the matrix factors is the value of the variable’s common or socialist contributions It is denoted by the symbol h_j whereas

$$h_j^2 = a_{j1}^2 + a_{j2}^2 + ... + a_{jm}^2$$

$$h_j^2 = \sum_{p=1}^{m} a_{jp}^2$$

Where the a_{jp} weight of the factor p is relative to the variable j, it is the coefficients of the matrix of factors, known as the saturation of the factors.

3 -Saturation Loading

Correlation coefficient or the variance between the variable or phrase and the component factor.

The total variance of the variable j can be represented as: $h_i^2 + s_i^2 + e_i^2 = 1$

And the ratio of the variance approved for the variable is the sum of the ratios of the varied commons as follows

$$R_{jj} = h_j^2 + s_j^2 = 1 - e_j^2$$

Where R_{jj} represent the approved transactions. One of the characteristics h_j is that they are positive between zero and the correct one meaning $0 \le h_j^2 \le 1$ (Feryal, 2006)

4-Rotation

An engineering method whose purpose is to make the larger loadings smaller and smaller saturation smaller than they were before rotation, as it reduces negative saturation and increases zero saturation in cases where there is no logical explanation for the negative signal of saturation. One of the most widely used recycling methods is Vairmax. (Saad, p. 13(

5- Kaiser-Mayer-Ulken test

To calculate the adequacy of the sample and to test it if the partial correlations between the variables are small 'and the value of this test ranges from (1-0) where the values close to (1) indicate the adequacy of the sample or it is appropriate' and values less than (0.5) indicate the insufficiency of the sample.

Bartlett's test

It is an indicator of the relationship between the variables and it must be statistically significant, which indicates that the matrix is the unit matrix (Abu Hashem, http: // ecsme.; Su.edu.sa(

Methods of factor analysis

1- Diagonal Method

This method is one of the straightforward and easy methods of factor analysis. It can be used in the event of few variables and leads to extracting the largest possible number of factors. This method requires previous and accurate knowledge of the values of the variability of variables. Without this knowledge it cannot be used. The diagonal method derives its name from being based directly on the use of diagonal values in the relational matrix. And you start extracting this value in its entirety in the first factor 'and thus the root of this value is the saturation of the first variable over the first factor' and it is called the radial saturation and so on (Hussam, p. 5)

2 -Central Method

The central method was the most widely used and popular method of global analysis until recently due to its ease of calculation as well as the conclusion of a few general factors', however this method lacks a number of important advantages' the most important of which is that it only extracts a limited amount of correlation variance '. The relational matrix, according to inaccurate estimates, uses the maximum correlation between the variable and any variable in the matrix, which is a measure that leads to demotion of the matrix (Hussam 2017(

3- Principal Components Method

The basic components method is one of the most important methods of factor analysis and comes at the forefront of simplicity. This analysis aims to take a variable p $x_1, x_2, ..., x_p$ and find a synthesis of these variables $y_1, y_2 ..., y_p$ to produce unrelated indicators. Weak association is a very useful feature because it indicates that indicators measure different dimensions of data. These indicators are y_1 arranged so that it has the most amount of variance y_2 'has the second largest amount of variance ... Thus, that is, where $\text{var}(y_1) \geq \text{var}(y_2) \geq ... \geq \text{var}(y_p)$ it denotes the variance $\text{var}(y_i)$ of the variable y_i for the studied data set. They are called y_i basic ingredients. And there is always hope when conducting an analysis of the basic components that the variations of most indicators are very few so that they can be neglected 'In this case the change in the group of variations can be described well by a few variables y with variations that cannot be ignored' and therefore we get a degree of economy in the calculations x_p as Dispersion in the original variable with fewer variables. [Faraj and Khalil, p. 71]

The main component is a linear combination of the original variables i.e. $x_1, x_2, ..., x_p$

$$Y_1 = a_{11}x_1 + a_{12}x_2 + ... + a_{1p}x_p$$

$$Y_2 = a_{21}x_1 + a_{22}x_2 + ... + a_{2p}x_p$$
$$\vdots$$
$$Y_i = a_{i1}x_1 + a_{i2}x_2 + ... + a_{ip}x_p$$

$$Y_i = \sum_{i=1}^{p} a_{ij}x_i \quad , \quad i, j = 1,2,..., p$$

Y_i It represents the primary component a_{ij} . It represents the coefficient of the variable j from component i, which represents the values of the distinct vectors accompanying the distinct roots (λ_i) of the array used. Each column of the array contains parameters of one of the basic a_{ij} components

$$\sum_{i=1}^{p} a_{ij}^2 = 1$$

And $Y_1, Y_2, ..., Y_p$ independent (not connected).

The base component Y_i variance is the distinct root (λ_i) accompanying the accompanying matrix $V(Y_i) = \lambda_i$

and surely $\text{var}(y_1) \geq \text{var}(y_2) \geq ... \geq \text{var}(y_p)$

And also that $\lambda_1 > \lambda_2 > ... > \lambda_p$. The total variance of the fundamental component is equal to the sum of the subjective values. The important characteristic of the distinct values is its sum equal to the sum of the diagonal elements of the correlation matrix or the co-variance matrix 'meaning $\sum_{i=1}^{p} \lambda_i = \sum_{i=1}^{p} a_{ii}$

That is, the sum of the variations of the basic compounds equals the sum of the variances of the primary variables (Abu Fakhida, p. 64). Linear structures coefficients must be chosen such that the standard vector values are i.e.

That all distinct roots must be positive, and their sum equals the sum of the diagonal elements of the matrix. The importance of the main component is estimated by the

amount of variation it explains relative to the total variance, i.e. the importance of the main component is. $\frac{\lambda_i}{\sum \lambda_i}$

The third topic: the applied side

Statistical method was applied to multivariate (factor analysis). Study of the variables that lead to an increase in the number of traffic accidents. The reasons leading to an increase in the number of traffic accidents of all kinds, collision and coup accidents and run over accidents were taken. My agencies:

Accidents by road type: x1: highway. X2: main road. x3: Subway. x4: Country road.

Number of accidents according to the time of the accident (photophobia): x5: Sunrise time. x6: sunset time. x7: daytime x8: night.

Number of accidents by accident: x9: road. X10: car x11: driver x12: pedestrian.

Number of accidents by type: x13: fatal. X14: fatal with injured. X15: injured only. X16: no injuries. The research included ((15 governorates other than the northern governorates due to the presence of data in the source (Nineveh, Salah al-Din, Kirkuk, Diyala, Anbar, Baghdad, Babil, Karbala, Najaf, Qadisiyah, Muthanna, Dhi Qar, Wasit, Maysan, Basra)

*** Results**

The first part of the results represents the R-Matrix correlation matrix shown in Table 1 and it is an identical matrix meaning that the top part of the main diameter is exactly the same as the part under the main diameter. The main diameter elements are equal to one, because each represents the value of the correlation coefficient between the variable and itself. Also, this matrix is anomalous with its determinant equal to zero. There is also a correlation between the study variables.

Table No. (1) shows the correlation matrix

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	x16
x1	1.000	.296	.033	-.018-	.489	.086	.496	.719	.652	.399	.416	.306	.707	-.146-	.520	-.084-
x2	.298	1.000	.431	.308	.504	.430	.899	.697	.172	.509	.932	.172	.557	.674	.770	.746
x3	.033	.431	1.000	.732	.589	.884	.589	.458	.473	.591	.629	.825	.646	.437	.790	.088
x4	-.018-	.308	.732	1.000	.747	.783	.345	.418	.452	.390	.457	.778	.445	.536	.617	-.007-
x5	.489	.504	.589	.747	1.000	.770	.522	.717	.712	.600	.611	.780	.671	.438	.770	.185
x6	.086	.430	.884	.783	.770	1.000	.455	.480	.502	.692	.563	.865	.587	.446	.740	.204
x7	.496	.899	.589	.345	.522	.455	1.000	.772	.402	.535	.968	.366	.794	.506	.903	.456
x8	.719	.697	.458	.418	.717	.480	.772	1.000	.575	.690	.777	.500	.771	.430	.844	.258
x9	.652	.172	.473	.452	.712	.502	.402	.575	1.000	.477	.332	.778	.781	.037	.590	-.193-
x10	.399	.509	.591	.390	.600	.692	.535	.690	.477	1.000	.523	.639	.621	.329	.682	.320
x11	.416	.932	.629	.457	.611	.563	.968	.777	.332	.523	1.000	.391	.734	.602	.922	.521
x12	.306	.172	.825	.778	.780	.865	.366	.500	.778	.639	.391	1.000	.706	.193	.698	-.197-
x13	.707	.557	.646	.445	.671	.587	.794	.771	.781	.621	.734	.706	1.000	.238	.873	.002
x14	-.146-	.674	.437	.536	.438	.446	.506	.430	.037	.329	.602	.193	.238	1.000	.478	.497
x15	.520	.770	.790	.617	.770	.740	.903	.844	.590	.682	.922	.698	.873	.478	1.000	.280
x16	-.084-	.746	.088	-.007-	.185	.204	.456	.258	-.193-	.320	.521	-.197-	.002	.497	.280	1.000

The second part of the results shown in Table 2) (represents the initial and extracted values for the Communalities' (note that the common factors have a percentage of the variance of the variables' as the lowest percentage is 0.572)) for the variable x_{10} that represents the car variable 'and the highest proportion of the variance of the variable is 0.989 for the variable x_{12} Which represents the infantry variable.

	Initial	Extraction
x1	1.000	.955
x2	1.000	.981
x3	1.000	.841
x4	1.000	.836
x5	1.000	.774
x6	1.000	.907
x7	1.000	.906
x8	1.000	.854
x9	1.000	.829
x10	1.000	.572
x11	1.000	.937
x12	1.000	.989
x13	1.000	.905
x14	1.000	.722
x15	1.000	.959
x16	1.000	.764

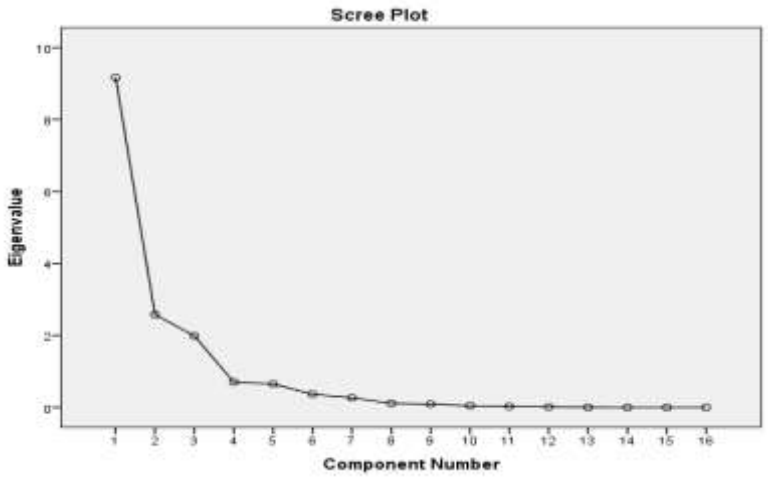
The third part of the results shown in Table No. (3), which shows the latent roots, contrast ratio and aggregate variance of the basic components. The data was reduced to three (basic factors), basic components that affect the increase in the number of car accidents, which represent the number of factors that exceed their correct one. These three factors constitute 85.821) of the total variance of the variables, and these factors have 57.279) for the first factor, (16.125) for the second factor, and (12.417) for the third factor.

Table No. (3)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	9.165	57.279	57.279	9.165	57.279	57.279	5.200	32.497	32.497
2	2.580	16.125	73.405	2.580	16.125	73.405	4.383	27.394	59.892
3	1.987	12.417	85.821	1.987	12.417	85.821	4.149	25.929	85.821
4	.703	4.394	90.215						
5	.649	4.057	94.272						
6	.366	2.287	96.559						
7	.264	1.651	98.210						
8	.112	.702	98.912						
9	.090	.563	99.475						
10	.043	.269	99.744						
11	.026	.161	99.905						
12	.014	.087	99.992						
13	.001	.008	100.000						
14	1.681E-005	.000	100.000						
15	1.002E-013	1.011E-013	100.000						
16	-1.000E-013	-1.001E-013	100.000						

Figure No. ((1 below shows that the number of basic factors that increase the value of the latent root from the correct one is (3) and the number of factors whose latent root is less than the correct one is 13)).

Figure No. (1)



The fourth part of the results, the Component Matrix, which represents Loading, is the first three factors that were extracted. The most powerful variable associated with the first factor is the variable x_{15} that represents only the type of accident injured 'where the saturation of the variable with the first component is (0.975). It is followed by the two variables x_{11} , x_{13} that represent the driver and fatal accidents' with a saturation of

the first component (0.870). And that the least saturated variables with the first component x_{16} is the one that represents a variable, there are no injuries as a result of the accident. As for the second component, its strongest association with the variable x_{16} reached saturation of its amount ((0.813 then the variable x_2 that represents the variable (rural road) with saturation amount (0.490) then the less saturated variables In the first component, x_{10} it represents the car with saturation of -0.039), but in the opposite direction.

As for the third component, its strongest correlation with the variable x_1 that represents the highway with saturation is (-0.814) and in the opposite direction. Then with the variable x_4 that represents the rural road variable with saturation amount (0.549) and the variable x_{14} representing the variable with a fatal accident with wounded with saturation amount of 0.433)).

Table No. (4)

	Component		
	1	2	3
x1	.502	-.202-	-.814-
x2	.747	.647	-.078-
x3	.793	-.179-	.424
x4	.684	-.265-	.546
x5	.852	-.211-	.065
x6	.807	-.210-	.459
x7	.841	.377	-.236-
x8	.851	.090	-.350-
x9	.666	-.558-	-.272-
x10	.755	-.039-	-.012-
x11	.870	.415	-.090-
x12	.765	-.591-	.233
x13	.870	-.202-	-.327-
x14	.543	.490	.433
x15	.975	.055	-.068-
x16	.302	.813	.105

The fifth part of the results represents the matrix of components after rotation, the analysis factors for each variable on each factor using Varimatrix with Kaiser shown in Table No. (5). From Table No. (5) above it can be concluded that the first factor and the variables that saturate it are. $(x_{12}, x_6, x_4, x_3, x_5, x_{15}, x_9, x_{10})$

The second factor and the variables that saturate it is (x_1, x_{13}, x_8, x_7)

The third factor and the variables that saturate it is $(x_2, x_{16}, x_{11,} x_{14})$

It is noted that the first factor that has great importance in influencing the increase in the number of car accidents in Iraq, which has the largest percentage of the total variance (57.279), then the second factor that got the second largest proportion of the total variation 16.125)) and finally the third factor that has a percentage of Total contrast (12,417).

Table No. (5)

	Component		
	1	2	3
x1	.502	-.202-	-.814-
x2	.747	.647	-.078-
x3	.793	-.179-	.424
x4	.684	-.265-	.546
x5	.852	-.211-	.065
x6	.807	-.210-	.459
x7	.841	.377	-.236-
x8	.851	.090	-.350-
x9	.666	-.558-	-.272-
x10	.755	-.039-	-.012-
x11	.870	.415	-.090-
x12	.765	-.591-	.233
x13	.870	-.202-	-.327-
x14	.543	.490	.433
x15	.975	.055	-.068-
x16	.302	.813	.105

• Conclusions through the statistical analysis of the study variables, it became clear the following

1 extracting three basic factors that make up(85,821) from the total variance of the study variables.

2) basic ingredients (x_3, x_2, x_1) It has a contrast ratio $(12.412, 16.125, 57.279)$

(3) The first factor is the inclusion of variables (pedestrians, sunset times, rural roads, sunrise times, sub roads, sunrise times, cars, roads, and pedestrians). That is, these variables are the primary causes of increased car accidents in which only the injured.

4) The second factor is the inclusion of variables (the highway, fatal accidents, night time and day time), meaning that these variables come second to the number of accidents, meaning that fatal accidents abound on the highway because of the speed.

5) The third factor is the inclusion of variables (the main road, no injuries, driver, and the presence of wounded with dead).

6) In the correlation matrix, the strongest correlation coefficient was between the driver variable and daylight time, meaning that the number of accidents caused by the driver occurs during the day, which causes the number of accidents to increase.

7- Recommendations The researchers recommend the following

1) The use of factor analysis in the study of multivariate phenomena for the ability of this method to reduce the number of variables, thus facilitating the study of these phenomena.
2) The necessity of providing the necessary lighting at night in the streets.
3) Putting speed cameras in highways.
4) Paying attention to the media aspect to educate citizens about traffic safety indicators.
5) Provide sufficient traffic lights at intersections.
6) Imposing stiff penalties for violators of traffic lights.

•Sources

-1 Abu Fakhidha, Ismail Radi, "Using Statistical Methods to Study the Role of Children in the Home Environment in Ramallah and Al-Bireh Governorate, Palestine", Beirut University, Birzeit College, pp. 63-64, 2013.

-2 Abu Hashem, Dr. Mohamed "Factor Analysis", King Saud University, College of Education, Department of Psychology, http://ecsme.ksu.edu.sa, entry date 15-12-2019.

-3 Hossam 'Saif (2017)" Using Factorial Analysis to Determine the Most Important Factors Affecting Iraqi Youth Migration "A Statistical Study of the Status of Youth Migration in Al-Qadisiyah" Al-Qadisiyah Journal of Computer Science and Mathematics "Volume (9) No. (1), pp. 1-11.

Khalaf, Nihad Sharif. (2012) "Using factor analysis to study the most important factors affecting child health in Salah al-Din Governorate. Tikrit Journal of Pure Sciences Volume 17' No. 1, p. 305.

-5 Saad, Dr. Alia Othman Sayed Saad. (2016) "Factor analysis of the causes of academic failure for Dammam University students - Saudi Arabia," Journal of Graduate Studies - Al-Neelain University, Volume 5, No. 15, pp. 2-25.

-6 Abdel-Razek, Khawla Bahgat 0 (2010). "The use of factor analysis to indicate the factors affecting the increase in the phenomenon of cheating among students." The Tech Magazine, Volume 11, Issue 9.

7 - Faraj 'Muhammad Hassan Mahmoud and Khalil', Faiza Muhammad Al Hassan (2014), "Using Factor Analysis to Determine the Most Important Factors Affecting the Emigration of Sudanese Scientific Competencies: A Statistical Economic Study of the Status of Sudanese Scientific Competencies in the Kingdom of Saudi Arabia" Amarabac Magazine 'Volume (5) No. (12 pp. 65-84.

Kazem, Feryal Mahmoud, "A statistical analysis of the Millennium Development Goals using the two methods of factor analysis and cluster analysis" Arab Institute for Training and Statistical Research, p. 23, 2006.

9- Nameg, Dr. Faisal Naji, “Employing Exploratory Factor Analysis to Extract Factors of Organizational Loyalty and Job Satisfaction” A Case Study at the Technical College of Management - Baghdad, Journal of Economic and Administrative Sciences, University of Baghdad, Volume 19, No. 20, pp. 96-123, 2013.

1- Morrison, D.F. "Multivariate Method, 2nd, Edition, McGraw, Hill, Moscer, C.I.1978.