Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.6, No.2, 2016

Using Factor Analysis to Study the Effecting Factor on Traffic Accidents

Layla A. Ahmed

Department of Mathematics, College of Education, University of Garmian, Kurdistan Region –Iraq

Abstract

This paper is concerned with the study of the factors that have significant effects on contracting traffic accidents. Another aim is to know whether the influence of each variable is independent or has a relation to that of other once. The sample of the research included (150) traffic accidents, the sample was collected from (Directorate of Traffic / Garmian) in the period (2013-2014). Measures of nine variables have been taken. And principal component method was used on the studied variables data to specify the importance of these variables. As well as, Varimax method was used to rotate the axis to get an easier and more specific result. The results have showed that the following variables have clear influences but their importance is different in terms of influencing on traffic accidents. We find that the variables (driving license) and (type of accident) have comes in first rank, while the other factors comes in later rank.

Keywords: Traffic accidents, Factor analysis, Principal component method, Rotation axes

1.1. Introduction

Traffic accidents are considered the most important types of accidents occurring in the country, which has become necessary to work to find solutions and suggestions to them. The statistical analysis of factor that influences the traffic accidents was carried out by using factor analysis method.

Factor analysis is a branch of multivariate analysis procedure that attempts to identify any und relying "factors" that are responsible for co variation among group independent variables. The goals of a factor analysis are typically to reduce the number of variables used to explain a relationship or to determine which variables show a relationship [9].

Factor analysis originated in psychological theory. Based on the work under taken by Pearson (1901) in which he proposed a" method of principal axes", Spearman (1904) began research on the general and specific factors of intelligence [14]. The term factor analysis was first introduced by Thurston (1931)[9]. Lewbel (1991) and Donal (1997) used the rank of a matrix to test for the number of factors, but these theories assume either N or T (the cross-section dimension and the time dimension, respectively) is fixed. Forni, Hallin, Lippi and Reichlin (2000) suggested a multivariate variant of the Akaike information criterion (AIC) but neither the theoretical nor the empirical properties of the criterion are known [3].

This study aims at determining the factors that have significant effects on traffic accidents. Another aim is to know whether the influence of each variable is independent or has a relation to that of other once.

1.2. Factor Analysis

The factor analysis model expresses each variable as a linear combination of underlying common factors $f_1, f_2, \dots\dots, f_m$, with an a companying error term to account for that part of the variable that is unique, the model is as follows [2][12][13]:

$$\underline{X}_{p\times n} = \underline{\mu}_{p\times 1} + A_{p\times m}\underline{F}_{m\times 1} + \underline{U}_{p\times 1} \quad \dots\dots (1)$$

Where:

m: The number of common factors (m<p).

A: Loading of the jth variable on the factor.

F: Common factors.

U: Specific factors.

μ: Mean of variables.

In factor analysis we begin with a set of variables $x_1, x_2, \dots\dots, x_k$.

These variables are usually standardized so that their variances are each equal to one and their covariance are correlation coefficients [8]. Assume that each x_i is a standardized variable,

$$x_i = \frac{(x_i - \bar{x}_i)}{s_i} \quad \dots\dots (2)$$

$$E(\underline{x}) = \underline{\mu} = \underline{0} \quad , \quad V(\underline{x}) = \underline{I}$$

Model (1) can be written:

$$\underline{X} = A\underline{F} + \underline{U} \quad \dots\dots (3)$$

The random vectors F and U are unobservable and uncorrelated.

$$E\begin{pmatrix}\underline{F}\\ \underline{U}\end{pmatrix}\begin{pmatrix}\underline{F} & \underline{U}\end{pmatrix} = \begin{bmatrix} E(\underline{F}\ \underline{F}) & E(\underline{F}\ \underline{U}) \\ E(\underline{U}\ \underline{F}) & E(\underline{U}\ \underline{U}) \end{bmatrix} = \begin{bmatrix} \Phi_{m\times m} & 0_{m\times p} \\ 0_{p\times m} & \Psi_{p\times p} \end{bmatrix} \quad \dots\dots (4)$$

Where:

Φ : Symmetric matrix of factor variance and covariance.

Ψ : Diagonal matrix of unique factor variances.

Thus the covariance of x can be written as:

$$E(X\acute{X}) = \Sigma_{P\times P} \quad \dots\dots (5)$$

Where Σ is a $P \times P$ population covariance matrix.

$$\Sigma = \mathrm{E}\big(A\underline{F} + \underline{U}\big)\big(A\underline{F} + \underline{U}\big)$$

$$\Sigma = A\mathrm{E}(\underline{F}\acute{F})\acute{A} + AE(\underline{F}\acute{U}) + E(\underline{U}\,\acute{F})\acute{A} + E(\underline{U}\,\underline{U}') \quad \text{...... (6)}$$

Since

$$\mathrm{E}(\underline{F}\acute{F}) = \Phi$$

$$E(\underline{F}\acute{U}) = E(\underline{U}\,\acute{F}) = 0$$

$$E(\underline{U}\,\underline{U}') = \Psi$$

Therefore $\Sigma = A\Phi\acute{A} + \Psi$ ...... (7)

1.3. Basic Assumptions of Factor Analysis

In factor analysis, we group variables by their correlations, such that variables in a group (factor) have high correlations with each other. Thus, for the purposes of factor analysis, it is important to understand how much of a variables variance is shared with other variables in that factor versus what cannot be shared. The total variance of any variable can be partitioned in to three types of variance [4]:

a. Common variance: Is defined as that variance in a variable that is shared with all other variables in the analysis, denoted by h_j^2.

$$h_j^2 = a_{j1}^2 + a_{j2}^2 + a_{j3}^2 + \cdots .. + a_{jm}^2 \quad \text{....... (8)}$$

b. Specific variance (also known as unique variance) is that the variance associated with only a specific variable. This variance cannot be explained by the correlations to the other variables but is still associated uniquely with a single variable.

$$u_j^2 = b_j^2 + e_j^2 \quad \text{......(9)}$$

Where:

u_j^2: Specific variance.

b_j^2: Special variance to variable j.

e_j^2: Error variance.

c. Error variance is also variance that cannot be explained by correlations with other variables, but it is due to unreliability in data gathering process, measurement error, or a random component in the measured phenomenon, denoted by e_j^2.

$$e_j^2 = 1 - (h_j^2 + b_j^2) \quad (10)$$

1.4. Commonalties

Is the proportion of the variance of an item that is accounted for by the common factors in a factor analysis, denoted by h_j^2.

$$h_j^2 = a_{j1}^2 + a_{j2}^2 + a_{j3}^2 + \cdots .. + a_{jm}^2$$

$$h_j^2 = \sum_{i=1}^{m} a_{ij}^2 , \quad \begin{cases} j = 1,2, \ldots ., p \\ i = 1,2, \ldots .., m \end{cases} \qquad \ldots\ldots (11)$$

$$0 \le h_j^2 \le 1$$

Where a_{ip}^2 represent the weight factor p for variable j.

1.5. Eigen value

The standardized variance associated with a particular factor. The sum of the eigen values cannot exceed the number of items in the analysis, since each item contributes 1 to the sum of variances [1]. An eigen vector of the matrix A as a vector u that satisfies the following equation [6]:

$$Au = \lambda u \qquad \ldots\ldots (12)$$

When rewritten, the equation becomes:

$$(A - \lambda I)u = 0 \qquad \ldots\ldots (13)$$

Where λ a scalar is called the eigen value associated to the eigenvector.

1.6. Principal Component Method

Principal component is considered the most important stages in the factor analysis method, and working to transform the variables associated to the new variables uncorrelated with each other.

The components are linear combinations weighted sums of the original variables [13].

$$Z_i = PC_i = a_{1i}X_1 + a_{2i}X_2 + \ldots \ldots + a_{pi}X_p \qquad \ldots\ldots (14)$$

$$PC_i = \sum_{j=1}^{p} a_{ji}X_j \ , \quad i, j = 1,2, \ldots .., p \qquad \ldots\ldots (15)$$

$$S = CD\acute{C} \qquad \ldots\ldots (16)$$

Where S is a sample covariance matrix and C is an orthogonal matrix constructed with normalized eigenvectors $(\acute{c}_i c_i = 1)$ of S as columns and D is a diagonal matrix with the eigenvalues $\lambda_1, \lambda_2, \ldots \ldots, \lambda_p$ of S on the diagonal:

$$D = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_p \end{pmatrix} \qquad \ldots (17)$$

$$D = D^{1/2}D^{1/2}$$

$$S = CD\acute{C} = CD^{1/2}D^{1/2}\,\acute{C}$$

$$= \left(CD^{1/2}\right)(C\,D^{1/2}\,\acute{)} \qquad \ldots\ldots (18)$$

This is of the form $S = A\acute{A}$, but we do not define A to be $CD^{1/2}$ because $CD^{1/2}$ is $P \times P$, and we are seeking a A that is $p \times m$ with $m < p$.We therefore define $D_1 = diag(\lambda_1, \lambda_2, \ldots .., \lambda_m)$ with the m largest eigenvalues $(\lambda_1 > \lambda_2 > \cdots . > \lambda_m)$ and $C_1 = (c_1, c_2, \ldots .., c_m)$ containing the corresponding eigenvectors. We then estimate A by the first m columns of $CD^{1/2}$,

$A = C_1 D_1^{1/2} = (\sqrt{\lambda_1}c_1, \sqrt{\lambda_2}c_2, \ldots., \sqrt{\lambda_m}c_m)$ …… (19)

Where A is $p \times m$, C_1 is $p \times m$, and $D_1^{1/2}$ is $m \times m$.

The i^{th} diagonal element of AÁ is the sum of squares of the i^{th} row of

A, or $\acute{a}_i a_i = \sum_{j=1}^{m} a_{ij}^2$. Hence to complete the approximation of S in (16), we define

$\psi_i = s_{ii} - \sum_{j=1}^{m} a_{ij}^2$ …… (20)

And write $S = AA\acute{} + \psi$ …… (21)

$h_i^2 = \sum_{j=1}^{m} a_{ij}^2$ … (22)

Which is the sum of squares of the i^{th} row of A. the sum of squares of the j^{th} column of A is the i^{th} eigenvalue of S:

$\sum_{i=1}^{p} a_{ij}^2 = \sum_{i=1}^{p} (\sqrt{\lambda_j}\, c_{ij})^2$

$= \lambda_j \sum_{i=1}^{p} {c_{ij}}^2 = \lambda_j$ ……(23)

Since the normalized eigenvectors (columns of C) have length 1. By equations (20) and (22), the variance of the i^{th} variable is partitioned into a part due to the factors and a part due uniquely to the variable:

$s_{ii} = h_i^2 + \psi_i$

$= a_{j1}^2 + a_{j2}^2 + a_{j3}^2 + \cdots .. + a_{jm}^2 + \psi_i$ …… (24)

Thus the jth factor contributes a_{ij}^2 to s_{ii}. The contribution of the jth factor to the total sample variance,

$tr(s) = s_{11} + s_{22} + \cdots + s_{pp}$, is, therefore,

Variance due to jth factor $= \sum_{i=1}^{p} a_{ij}^2 = a_{1j}^2 + a_{2j}^2 + \cdots .. + a_{pj}^2$ ..… (25)

Therefore $\frac{\sum_{i=1}^{p} a_{ij}^2}{tr(S)} = \frac{\lambda_j}{tr(S)}$ …… (26)

We can use standardized variables and work with the correlation matrix R.

$\frac{\sum_{i=1}^{p} a_{ij}^2}{tr(R)} = \frac{\lambda_j}{p}$ …… (27)

Where p is the number of variables.

1.7. Rotation Axes

1.7.1. Simple Structure

Most of the rationale for rotating comes from Thurston (1947) and Cattell (1978) who defended its use because this procedure simplifies the factor structure and therefore makes interpretation easier and more reliable easier to replicate with different data samples [1],[11].Thurston (1947) first proposed and argued for five criteria that needed to be met for simple structure to be achieved[5]:

- **a.** Each variable should produce at least one zero loading on some factors.
- **b.** Each factor should have at least as many zero loadings as there are factors.
- **c.** Each pair of factors should have variables with significant loadings on one and zero loadings on the other.
- **d.** Each pair of factors should have a large proportion of zero loadings on both factors.
- **e.** Each pair of factors should have only a few complex variables.

1.7.2. Orthogonal Rotation and Oblique Rotation

An orthogonal rotation is specified by a rotation matrix denoted R, where the rows stand for the original factors and the columns for the new (rotated) factors. There are several methods for orthogonal rotation such as the varimax, Quartimax, Equimax and Orthomax [1],[11].

In oblique rotations the new axes are free to take any position in the factor space, but the degree of correlation allowed among factors is, in general, small because two highly correlated factors are better interpreted as only one factor. There are several methods for orthogonal rotation such as the Quartimin, Promax, Procrustes.

1.7.3. The Kaiser -Varimax Method

A popular scheme for rotation was suggested by Henry Kaiser in (1958). He produced a method for orthogonal rotation of factors, called the varimax rotation [7], achieved by maximizing the sum of the variances of the squared factor loadings within each factor [2].

1.8. Number of Factors

The Kaiser method proposed by Kaiser (1960) is perhaps the best know and most utilized in practice. According to this method, only the factors that have eigen values greater than one are retained for interpretation [10].

2. Data Analysis and Results

2.1. Data Description

The data that were used in this research is data from the statistical report of traffic accidents recorded from (Directorate of Traffic / Garmian) in the period (2013-2014). The extraction results of analyzes using the statistical program (SPSS V.22) includes a set of data variables:

X_1: Age ($< 30\ years = 1$, $\geq 30\ years = 2$)
X_2: Type of composite (car) (small car (taxi) =1, Bus =2, Lorry=3)

X_3: Type of accidents (Coup =1, Collision = 2, Run over =3)

X_4: Accident time (Day =1, Night =2)

X_5: Weather conditions (Rainy =1, Sunny =2, Cloudy =3)

X_6: The place of the accident (Inside the city =1, Outside the city = 2)

X_7: Driving license (Yes =1, No =2)

X_8: Due to the drinking (Yes =1, No =2)

X_9: The cause of traffic accidents (Excess speed =1, Passing wrong =2)

2.2. Analysis of the Results

After analyzing the correlation matrix by principal component method this method is the most widely used for determining a first set of the loadings and seeks values of the loadings that bring the estimate of the total communality as close as possible to the total of the observed variances, it became clear the existence of five factors represented by the number of egien values that greater than one. Where the extraction factors that accounted for 65.978% of the total variance of the variables as shown in table (1).

Table (1) Initial Eigen values and Rotation Sums of Squared Loadings

	Initial Eigen values			Rotation Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
Age	1.439	15.992	15.992	1.300	14.447	14.447
Type of composite	1.268	14.091	30.083	1.262	14.018	28.465
Type of accident	1.122	12.465	42.548	1.183	13.144	41.609
Time of accident	1.090	12.115	54.663	1.113	12.371	53.980
Weather conditions	1.018	11.315	65.978	1.080	11.998	65.978
place of accident	.954	10.600	76.578			
Driving license	.829	9.207	85.786			
Due to the drinking	.692	7.684	93.470			
Cause of accident	.588	6.530	100.000			

The results in table (2), refers to the components matrix which represents the results of extraction factors after rotation calculated according the method of Varimax with Kaisers normalization.

Table (2) Rotated Component Matrix

	Rotated Component Matrix[a] Component				
Component	1	2	3	4	5
Age	.126	.357	-.052	-.356	-.048
Type of composite	.016	-.106	.864	-.032	.026
Type of accident	.596	.140	.491	-.071	-.108
Time of accident	.207	-.070	-.169	.820	.067
Weather conditions	-.046	-.105	-.041	-.005	.869
place of accident	-.002	-.788	.148	.042	.176
Driving license	.868	.009	-.037	.073	-.009
Due to the drinking	.009	.609	.275	.109	.471
Cause of accident	-.362	.308	.253	.538	-.231

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
Rotation converged in 9 iterations.

Conclusions

The results have showed that the following factors that accounted for 65.978 % of the total variance of the variables have clear influences but their importance is different in terms of influencing on traffic accidents.

1. The First Factor

This factor has a great importance in influencing road accidents where he explains (14.447) of the total variance, included two variables have the greatest impact on this factor which are driving license, type of accident with components (0.868, 0.596) respectively.

2. The Second Factor

This factor ranked second in terms of importance in the interpretation of the relationship between the variables, where he explains (14.018) of the total variance, this factor contains about three variables that includes: due to the drinking, the place of the accident and age, with components (0.609, -0.788, 0.357) respectively.

3. The Third Factor

This factor ranked third in terms of importance in influencing road accidents, where he explains (13.144) of the total variance, this factor contains only one variable type of composite, with component (0.864).

4. The Fourth Factor

This factor ranked fourth in terms of importance in the interpretation of the relationship between the variables, where he explains (12.371) of the total variance, this factor contains about two variables which contains: accident time and the cause of traffic accidents, with components (0.82, 0.538) respectively.

5. The Fifth Factor

This factor ranked fifth and last in terms of importance in the importance in influencing road accidents, where he explains (11.998) of the total variance, this factor contains one variable weather conditions, with its components (0.869).

Recommendations

1. Militancy in giving licenses leadership and increased attention to verify the fitness standards for drivers and accuracy of the medical examination process for those who want to get driving licenses.
2. Militancy in the activation of the traffic lows and increase the number of speed cameras.
3. Enter reporter traffic culture in different academic levels commensurate with each stage.

References

[1] Abdi, H. and Williams, l. j.,(2010), Principal Component Analysis, Wiley Interdisciplinary Reviews, Computational Statistics, Vol.2, pp1-47.

[2] Afifi, A., Clark, V. A. and Susanne M.,(2004), Computer- Aided Multivariate Analysis, 4th Edition, Chapman & Hall/CRC, London, New York.

[3] Bai, J. and Serena, N., (2002), Determining the Number of Factor in Approximate Factor Models, Journal of Econometrica, Vol.70, No.1, pp 191-221.

[4] Brace, N., Kemp, R. and Snelgar, R.,(2006), SPSS for Psychologists, Versions 12and 13, 3rd Edition, Palgrave Macmillan, New York.

[5] Brown, J. D., (2009), Choosing the Right Type of Rotation in PCA and EFA, JALT Testing and Evaluation SIG Newsletter, Vol.13, No.3, pp 20-25.

[6] Coombe, G., (2006), An Introduction to Principal Component Analysis and Online Singular Value Decomposition.

[7] Hair, J. F., and others, (2006), Multivariate Data Analysis, 6th Edition, Pearson prentice Hall, Upper Saddle River, Inc. [8] Hardle,W. and Simar, L., (2003), Applied Multivariate Statistical Analysis, Method and Technologies, Version 29.

[9] Lawley, D. N. and Maxwell, A. E., (1962), Factor Analysis as a Statistical Method, Journal of the Royal Statistical Society, Vol.12, No.3, pp209-229.

[10] Ledesma, R. D. and Valero- Mora, P., (2007), Determining the Number of Factor to Retain in EFA: an easy to Use Computer Program for Carrying out Parallel Analysis Practical Assessment, Journal of Research and Evaluation, Vol. 12, No. 2.

[11] Mulaik, S. A., (1972), The Foundations of Factor Analysis, McGraw- - Hill Book Company, New York.

[12] Preacher, K. J. and Zhang, G., (2013), Choosing the Optimal Number of Factors in Exploratory Factor Analysis: A Model Selection Perspective, Multivariate Behavioral Research, Vol. 48, pp28-56.

[13] Rencher, A. C., (2002), Methods of Multivariate Analysis, 2nd Edition, John Wiley and Sons Inc.