RESEARCH REVIEW JOURNALS
Volume-03
Issue-07
July-2018

ISSN: 2455-3085 (Online)
RESEARCH REVIEW International Journal of Multidisciplinary
www.rrjournals.com [UGC Listed Journal]

An application of Factor Analysis on GSRTC Data – A case study of Ahmedabad Depot

[*1]Bhavika D. Shah & [2]Pravender

[1]*Research Scholar, Department of Statistics, Gujarat University, Ahmedabad – 380009 (India)*
[2]*Research Guide, Department of Statistics, Gujarat University, Ahmedabad – 380009 (India)*

ARTICLE DETAILS

Article History
Published Online: 05 July 2018

Keywords
GSRTC, Bartlett's test, principle component analysis

[*]Corresponding Author
Email: bdshah64[at]yahoo.in

ABSTRACT

This research paper is working to test an application of factor analysis and formation of model of factor analysis. Finally, it is tested for GSRTC data sets for the year 2005 to 2015 for selected factors affecting to the margin of GSRTC. The factors under study are taken for sixteen depot of Gujarat state. The concluded results are determined by taking an application of Bartlett's test and principle component analysis.

1. Introduction

Factor analysis is a measurable model that permits clarifying the connections between are substantial quantities of observed relative factors through few uncorrelated undetectable components.

The starting point of factor analysis goes back to a work done by Spearman in 1904. Around then psychometricians were profoundly engaged with the endeavor to suitably evaluate human knowledge, and Spearman's work given an extremely sharp and valuable tool that is still at the bases of the most exceptional tool for estimating insight.

Knowledge is the model of an immense class of factors that are not directly observed characterized as inactive factors, but rather can be estimated in a circuitous path through the investigation of recognizable factors firmly connected to the inert ones. Idle factors are basic to numerous examination fields other than brain research, from prescription to hereditary qualities; frame back to financial aspects and this clarify the still clear enthusiasm towards Factor investigation.

Spearman considered the accompanying relationship framework between youngsters' examination execution in Local Language (X1), French Language (X2) and English Language (X3):

$$r = \begin{bmatrix} 1 & 0.84 & 0.75 \\ \square & 1 & 0.69 \\ \square & \square & 1 \end{bmatrix}$$

Spearman saw a high positive connection between the scores and hypothesized, that it was because of the relationship of the three observed factors with a further surreptitiously factor that he called knowledge or general capacity. On the off chance that his presumption was valid, than he expected that the fractional connection coefficients, figured between the observed factors subsequent to controlling for the normal inert one, would vanish.

Beginning from this instinct, he figured the accompanying model, which, as we will find in the accompanying, can flawlessly satisfy the objective:

$$X1 = \Delta 1f + U1, \quad X2 = \Delta 2f + U2, \quad X3 = \Delta 3f + U3$$

Where f is the basic factor Δ1, Δ2, Δ3 are the factor loading and U1, U2, U3 are the one of a kind or particular elements.

The factor loading show how much the basic factor adds to the diverse observational estimations of the x factors; the one of a kind variables speak to residuals, irregular clamor terms.The other than telling that an analysis just offers a surmised measure of the subject's capacity, likewise portray, for every person, how much his outcome on a given subject, say French, contrasts from his general capacity. Spearman's model can be summed up to incorporate in excess of one basic factor:

$$Xi = \Delta 1f1 + \Delta 2f2 + \cdots + \Delta ikfk + \cdots + \Delta infn + Ui$$

2. The Factor Model

Give X a chance to be a k-dimensional arbitrary vector with expected esteem μ and covariance framework ∑. Δm factor display for x holds on the off chance that it can be deteriorated as:

$$X = \Delta ifi + U + \mu$$

On the off chance that we accept to manage mean focused x factors at that point, with no misfortune in sweeping statement, the model will be X=Δifi + U .... Eq. (i)

Where Δ is the k × n factor stacking network; f is the n × 1 arbitrary vector of basic variables and U is the k × 1 irregular vector of one of a kind components.

The model resembles a straight relapse demonstrate, yet for this situation every components in right hand side of the equivalent sign are obscure. With a specific end goal to lessen indeterminacy, we can force the accompanying imperatives:

E (fi) = 0 and E (U) = 0

this condition is splendidly suitable with the way that the work with mean focused data

$E(fi \times fi^T) = 0$

It demonstrates that the basic components are institutionalized uncorrelated irregular factors: their changes are equivalent to 1 and their co-variance are 0. This suspicion could likewise be casual, while the accompanying ones are entirely required.

$E(Ui \times Ui^T) = \omega$

$$\text{Here, } \omega = \begin{bmatrix} \omega 11 & 0\ldots & \ldots & 0 \\ \cdots & \cdots\cdots & \cdots & \cdots \\ \cdots & \cdots\cdots & \cdots & \cdots \\ 0 & \ldots\omega ii & \ldots. & 0 \\ \cdots\cdots & \cdots\cdots & \cdots & \cdots \\ \cdots & \cdots\cdots & \cdots & \cdots \\ 0 & \ldots\ldots & \cdots & \omega nn \end{bmatrix}$$

It shows a diagonal matrix. This means that the unique factors are uncorrelated and may be have heteroskedasticity.

Therefore, $E(fiUi^T) = 0$ and $E(Uifi^T) = 0$ that means that the unique factors are uncorrelated with the commonfactors.

On the off chance that a factor display fulfilling the previously mentioned conditions holds, the covariance grid of the watched factors X can be decayed as takes after:

$$\begin{aligned} \Sigma &= E\,(XiXi^T) \\ &= E\,[(\Delta ifi + Ui)(\Delta ifi + Ui)^T] \\ &= E\,(\Delta ififi^T \Delta i^T + \Delta fi\, Ui^T + Uifi^T \Delta^T + Ui\, Ui^T] \\ &= \Delta E\,(fifi^T)\Delta^T + \Delta E\,(fi\, Ui^T) + E\,(Ui\, fi^T)\Delta^T \\ &\quad + E\,(Ui\, Ui^T) \end{aligned}$$

… Eq. (ii)

$$= \Delta I \Delta^T + \Delta 0 + 0\Delta^T + \omega$$
$$= \Delta\Delta^T + \omega$$

The inverse is likewise valid: if the covariance grid of the watched factors X can be deteriorated as in condition (ii) at that point the direct factor display (i) embraces.

As is corner-to-corner, deterioration (ii) plainly demonstrates that the regular factors represent all the watched co-variance and gives hypothetical motivation to Spearman's instinct.

It merits having a more critical take a gander at the corner-to-corner components of the frameworks on the two sides of the correspondence sign in deterioration (ii).

$$V(Xi) = s = \sum_{i=1}^{n} \theta_{ij}^2 + \gamma_{ij} = L_i^2 + \gamma_{ij}$$

The measure $\sum_{i=1}^{n} \theta_{ij}^2 = L_i^2$ presents communalities of factors; it shows the area of variance in Xi which is explained by the collective factors or concluded that it shows the variance of Xi distributed with the other factors. γ_{ij} is presenting unique variance, it shows the variance of the ith factor which is not accounted for by the collective factors.

From the suppositions on the normal and the novel factors that have been depicted a further intriguing portrayal for infers. How about consider the co-variance between the observed factors Xi and the normal variables f:

$$\begin{aligned} Cov(Xi, fi) &= E(Xifi^T) = E[(\Delta fi + Ui)fi^T] \\ &= \Delta E(fifi^T) + E(Uifi^T) = \Delta \end{aligned}$$

It clears that the factor loading matrix Δ is also the covariance matrix between X and f.

3. Rotation of Factors

To enhance interpretability of the factor loadings one can depend on the invariance to orthogonal turn property of the factor display. In 1947, Thurston gave a meaning of how an interpretable factor structure ought to be. The factors ought to be distinguishable into gatherings to such an extent that the loadings inside each gathering are high on a solitary factor, maybe direct to low on a couple of components and immaterial on the rest of the elements. One approach to get a factor-stacking lattice fulfilling such a condition is given by the supposed Varimax revolution. It searches for an orthogonal pivot of the factor-stacking grid, with the end goal that the accompanying foundation is amplified

$$S = \sum_{i=1}^{n} \left\{ \frac{\sum_{i=1}^{q} b_{ik}^4}{q} - \left(\frac{\sum_{i=1}^{q} b_{ik}^2}{q} \right)^2 \right\}$$

$$\text{Here, } b_{ik} = \frac{\theta_{ik}}{\sqrt{(\sum_{i=1}^{n} \theta_{ik}^2)}} = \frac{\theta_{ik}}{L_i}$$

It ought to be noticed that S is the aggregate of the differences of the squared normalized inside each column factor scores for each factor. Expanding it makes the substantial coefficients end up bigger and the little coefficients to approach 0.

4. Factor Scores

After the factor loadings and the unique variances have been estimated, we might be interested in estimating, for each statistical unit whose observed vector is *Xi*, the corresponding vector of factor scores fi. If for instance the first factor is intelligence, this could also allow us to rank the individuals according to their scores on this factor, from the most to the least intelligent ones. Two methods exist in popular use for factor score estimation.

4.1 Thompson Estimation

The strategy proposed by Thompson characterizes the factor scores as straight mixes of the watched factors limited the squared expected forecast mistake. For the kthfactor fk, the comparing gauge is given by $f_k^* = A^T X = X^T A_k$ where A_k is a q× 1 vector. According to Thompson's approach A_k should be chosen so that

$E(f_k^* - f_k)^2 = E\,(X^T A_k - f_k)^2$ is diminished. In the wake of separating regarding A_k and setting, the subordinates' equivalent to 0it is acquired:

$$\begin{aligned} &= E\,[2X(X^T A_k - f_k)] \\ &= 2[E\,(XX^T)A_k - E\,(Xf_k)] \\ &= 2(\sum A_k - \Delta_k) = 0 \end{aligned}$$

Here Δ_k is the kth column of Δ.

Thus, $A_k = \sum A_k - \Delta_k$ and $f_k^{*}=\Delta_k^T \sum X$. Than it can be$f^* = \Delta_k^T \sum X$. After using applications ofalgebra, a different expression for f^* can be:
$f^* = (I + \Delta_k^T \omega^{-1}\Delta)^{-1}\Delta^T \omega^{-1} X$. Both the measured tools are producing same results.

4.2 Bartlett Estimation

After the factor loading and the one of a kind differences have been evaluated, the factor model can be viewed as a direct multivariate relapse display where f is the obscure vector parameter and the residuals are uncorrelated however heteroskedasticity. Estimation can be tended to by weighted least squares.
An estimate of f is to be required to define,

$$U^T \omega^{-1} U = (X - \Delta f)^T \omega^{-1} (X - \Delta f)$$

ismust be minimum. After differentiating w.r.t. f and testing the first necessary condition(the first order derivatives equal to 0) it is found

$$-2\Delta^T \omega^{-1}(X - \Delta f)$$
$$= 2\Delta^T \omega^{-1}\Delta f - \Delta^T \omega^{-1} X = 0$$
$$\therefore f^* = (\Delta^T \omega^{-1} \Delta)^{-1}. \Delta^T \omega^{-1} X$$

Bartlett's estimator f^* is unbiased thus,

$$E(f^*/f) = E((\Delta^T \omega^{-1}\Delta)^{-1}. \Delta^T \omega^{-1} X/f)$$
$$= (\Delta^T \omega^{-1}\Delta)^{-1}. \Delta^T \omega^{-1} E(X/f)$$
$$= (\Delta^T \omega^{-1}\Delta)^{-1}. \Delta^T \omega^{-1} \Delta f$$
$$= f$$

The Bartlett's estimator has larger mean squared estimate error than estimator ofThompson.

5. Factor Analysis (FA) and Principle Component Analysis (PCA)

It merits finishing this part by focusing on the associations and the differences amongst FA and PCA.

The two strategies have the point of diminishing the dimensionality of a vector of irregular factors. Be that as it may, while FA accepts a model, PCA is only an information change and consequently, it generally exists.

Moreover, while FA goes for covariance or correlations, PCA just focuses on variances. In spite of these reasonable contrasts, there have been endeavors, for the most part previously, to utilize PCA with a specific end goal to gauge the factor display. In the accompanying, it can demonstrate that without a doubt PCA might be lacking when the objective of the exploration is fitting a factor testing.

Give X a chance to be the typical q dimensional arbitrary vector and Y the q dimensional vector of the relating chief segments$Y = \Delta i^T Xi$ with Δ theorthonormal matrix whose columns are the eigenvectors of the covariance matrix of the Xi given variables.

Because of the properties of Δ it will also be Xi = ΔiYi. Δcan be rotten into two represent matrices Δincovering the eigenvectors relating to the first n eigenvalues and Δi_{q-n}covering the enduring ones $\Delta i = (\Delta in/\Delta i_{q-n})$. The parallelmeasureis taken for vector Yi,
Thus,

$$Yi = \frac{Y_{in}}{Y_{iq-n}}$$
$$\therefore Xi = (\Delta in/\Delta i_{q-n}).\frac{Yi_n}{Yi_{q-n}}$$
$$= \Delta_{in} Y_{in} + \Delta i_{q-n} Y i_{q-n} \ldots . Eq.(iii)$$
$$= \Delta i_n . H_{in}^{1/2} H_{in}^{-1/2} Y i_n + \Delta i_{q-n} Y i_{q-n}$$

Here, H_n presents diagonal matrix of the n eigenvalues. Also, $\Delta i_n . Hi_n^{1/2} = \Delta i$ and $Hi_n^{-1/2} Y i_n = fi$ and $\Delta i_{q-n} Y i_{q-n} = \pi$, the derived function can be rewrite as $Xi = \Delta i fi + \pi$.

The introduced f factors have the similar properties as linear factor model. It can be presented as:

$$E(fi fi^T) = E\left(Hi_n^{-\frac{1}{2}} Y i_n . Y_{q-n}^T . \Delta_{q-n}^T\right)$$
$$= Hi_n^{-\frac{1}{2}} E\,(Y i_n . Y i_n^T)/Hi_n^{-\frac{1}{2}}$$
$$= Hi_n^{-\frac{1}{2}} Hi_n \, . \, Hi_n^{-\frac{1}{2}} = I$$

As the covariance matrix of n PCA is H_n also,

$$E(fi\pi^T) = E\,(Hi_n^{-\frac{1}{2}} Y i_n . Y i_{q-n}^T . \Delta i_{q-n}^T)$$
$$= Hi_n^{-\frac{1}{2}} E\,(Y i_n . Y i_{q-n}^T)\, \Delta i_{q-n}^T = 0$$

As the first n and the last q – n principle components are uncorrelated, the new unique factors π are correlated. The reverses to linear factor model molds rendering to which the associated factors entirely clarify the experiential covariance as:

$$= E\,(\pi\pi^T) = E(\Delta i_{q-n}/Y i_{q-n} \, . \, Y i^T{}_{q-n} . \Delta i^T{}_{q-n})$$
$$= \Delta_{q-n} E\left(Y i_{q-n} . Y i^T{}_{q-n}\right) \Delta i^T{}_{q-n}$$
$$= \Delta i_{q-n} . Hi_{q-n} . \Delta i^T{}_{q-n}$$

The Hi_{q-n} is the covariance matrix of the last q – n principle component and thus, it is diagonal; its diagonal elements are different and thus, $\Delta i_{q-n} . Hi_{q-n} . \Delta i^T{}_{q-n}$ is not diagonal.

6. Application to GSRTC Data

6.1 Theoretical Frame Work

The main objective of this study is to test the designed frame of factor analysis to test an application on GSRTC. The study is executed for selected sixteen depots of Gujarat Statefor financial year 2005 to 2016. The data are compiled in form of cross sectional time series – Panel structure. The defined parameters under construction of panel are defines as follows:

6.1.1 Effective KM (Eff)

The total load carried by bus to provide services to the passengers is called effective kilometers.

6.1.2 No. of Passengers (P)

Total number of passengers travelled during the year by taking services of GSRTC buses.

6.1.3 Total EPKM (EPKM)

Earning per Kilometer is the ratio between total revenue and effective kilometers.

6.1.4 Total CPKM (CPKM)

Cost per Kilometer is computed by taking ratio of total cost to effective kilometers.

6.1.5 Margin (Loss)

The margin is computed by subtracting the total cost from total earning. Positive margin values indicate the profit, while negative margin values indicate loss or deficit. Margin is calculated by following formula: Margin = Total Earnings – Total Cost

6.1.6 Load Factor Percentage (LF)

The load factor represent the percentage of seating capacity offered to seating capacity actually occupied.

6.1.7 Vehicle Utilized per day (VU)

Total running of vehicle travelled to complete shifts assigned to the crew.

6.1.8 Fleet Utilization (FU)

Fleet Utilization is a function, which allows GSRTC, to rely on transportation in business to remove or minimize the risks associated with vehicle investment, improving efficiency, productivity and reducing their overall transportation and staff costs, providing 100% compliance with government legislation.

6.1.9 Crew Utilization (CU)

Crew Utilization for GSRTC, otherwise known as crewing, are the services renderedhighbred by GSRTC for operating the services.

6.1.10 Diesel KMPL (D)

It shows usage of diesel in kilometer per liter during a year.

6.1.11 Engine Oil KMPL (EO)

It shows the usage of engine oil top up in kilometer per liter during a year.

6.1.12 Break Down (BD)

It shows until how many times a failure of the engine or other working parts of a vehicle or machine during a year.

6.1.13 Accidents (A)

Disaster type term used to describe technological transport accidents involving mechanized modes of transport.

6.2 Model Testing On GSRTC Data

6.2.1 Correlation Matrix

Table 1.1 Correlation Matrix

	Eff	P	EPKM	CPKM	Loss	LF	VU	FU	CU	D	EO	BD	A
Eff	1	**.84**	-.07	-.24	.46	.19	-.18	.04	-.12	.34	.423	.140	**.69**
P	**.84**	1	-.19	-.19	.49	.11	-.41	.002	-.40	.10	.249	.235	**.76**
EPKM	-.08	-.19	1	**.56**	.06	.11	.21	-.14	**.51**	.45	.025	-.29	-.48
CPKM	-.24	-.19	**.56**	1	.09	.011	.06	-.08	.21	.12	-.18	-.12	-.33
Loss	.46	.49	.07	.09	1	.09	-.29	-.02	-.29	.005	.097	.18	.37
LF	.19	.11	.12	.01	.09	1	.22	.10	.43	.46	**.58**	-.49	-.08
VU	-.19	-.41	.21	.06	-.29	.22	1	.01	**.61**	.38	.226	**-.50**	-.45
FU	.04	.002	-.14	-.08	-.02	.11	.014	1	-.04	.024	.087	.012	.066
CU	-.12	-.40	**.51**	.21	-.29	.43	**.61**	-.04	1	**.67**	.367	**-.72**	**-.59**
D	.34	.10	.45	.12	.05	.46	.39	.024	**.67**	1	**.71**	**-.66**	-.23
EO	.42	.25	.03	-.18	.097	.57	.23	.087	.37	.71	1	**-.56**	.063
BD	.14	.24	-.29	-.12	.18	-.49	**-.50**	.012	**-.72**	**-.66**	**-.56**	1	**.51**
A	**.69**	**.76**	-.48	-.33	.37	-.07	-.45	.066	**-.59**	-.22	.06	**.51**	1

Table 1.1 is presenting the correlation matrix between the selected factors under study. It can be seen from table 1.1 that the correlation between Effective Kilometers and Passengersis found 0.84. It cleared that as much the buses can run they will get higher number of passengers. Thus, the correlation between them is found higher. The correlation between total effective kilometers and accidentsis found0.69. It is also thinkable that higher running of buses may cause higher accidents. The correlation betweentotal number of passengerstravelled and number of accidentsis 0.76, shows higher correlation. The chances of accidents may arise if total number of passengers is higher. The relation between earning per kilometer and cost per kilometeris found 0.56, similarlythe relation between earning per kilometer and crew utilizationhaving correlation of 0.51. Load Factor per passengerconsume more Engine Oilthus the correlation among them is computed 0.58.The vehicle utilization andcrew utilizationare having correlation of 0.61.Total breakdown is negatively associated with vehicle utilization. Higher the usage of vehicle may cause to increase the totalbreakdowns it is found (-0.50), the correlation between crew utilization and usage of diesel are having correlation of 0.67, whereas it has negative relation with break downwhich is computed (-0.72). The usage of diesel andengine oilare having positive correlation, it is found0.71.The usage of diesel is negatively associated with total break down and is found(-0.66).

6.2.2 KMO- Bartlett's Test

The Bartlett's test looks at the observed relationship lattice to the character network. As such, it checks if there is a sure excess between the factors that can outline with a couple of number of variables. On the off chance that the factors are splendidly connected, just a single factor is adequate. In the event that they are orthogonal, we require the same number of components as factors. In this last case, the relationship network is the same as the character framework. A basic technique is to picture the relationship framework. In the event that the qualities outside the diagonal are regularly high (in total esteem), a few factors are corresponded; if most these qualities are close to zero. In general, the computed value of Kaiser Meyer – Olking value at least expected more than 0.50. Table 1.2 shows the computed value as 0.744.

Table 1.2 KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.744
Bartlett's Test of Sphericity	Approx. Chi-Square	1696.9
	df	78
	Sig.	.000

A few factors are corresponded. Here, the objective is just to get a general impression about the excess between the factors. It should affirm this with a thorough measurable technique.

Table 1.2 shows the Bartlett's test. It checks if the observed correlation matrix deviates meaningfully from the identity matrix i.e. H0: the parameters are orthogonal. The factor analysis can make a fare compression of the existing data only if reject the null hypothesis. In order to measure the overall relation between the variables, we compute the determinant of the correlation matrix |R|. Under H0, |R| = 1; if the variables are highly correlated, we have |R| ≈ 0. The Bartlett's test statistic indicates to what extent it deviate from the reference situation |R| = 1. It can be computed as

$$\chi^2 = -\left[n - 1 - \frac{2p+5}{6}\right] \times ln|\mathrm{R}|$$

Under H_0, it follows a χ^2 distribution with a $\{p \times \frac{(p-1)}{2}\}$degree of freedom. The null hypothesis is rejected at 5% level of significant for application of principle component analysis. In general, the principle component analysis is advisable to use when the proportion of total observations under study and total observations having response lower than five.

The KMO measured has a similar objective. It checks on the off chance that we can factorize productively the first factors. Be that as it may, it depends on another thought. The relationship lattice is dependably the beginning stage. It is known that the factors are associated, thus the others can affect the association between two factors. Along these lines, utilize the fractional connection with a specific end goal to quantify the connection between two factors by expelling the impact of the rest of the factors. The KMO list analyzes the estimations of relation between factors.Table 1.2 shows that the KMO value is high (near to1), the factor analysis can act productively; if KMO is low (0), the PCA is applicable.

6.2.3 Communalities

Table 1.3 Communalities

	Initial	Extraction
Effective KM	1.000	.850
No. of Passengers	1.000	.866
Total EPKM	1.000	.793
Total CPKM	1.000	.700
Loss	1.000	.588
Load Factor %	1.000	.520
Vehicle Utilized per day	1.000	.533
Fleet Utilization	**1.000**	**.115**
Crew Utilization	1.000	.824
Diesel KMPL	1.000	.831
Engine Oil KMPL	1.000	.798
Break Down	1.000	.767
Accidents	1.000	.843
Extraction Method: Principal Component Analysis.		

A. Communalities – This is the extent of every factor's fluctuation that can be clarified by the elements. It is additionally noted as sum of square of factors and can be characterized as the entirety of squared factor loadings for the factors.

B. Initial – With key factor pivot considering, the underlying qualities on the corner to corner of the relationship grid are controlled by the squared different connection of the variable with alternate factors. Table 1.3 is computed for perfect co-efficient values of each parameter hence it do not have corner to corner value.

C. Extraction – The value in this column demonstrate the extent of every factor's change that can be clarified by the held components. Factors with high values are all around spoke in the normal factor space, while factors with low values are not very much spoke. Table 1.3 shows that while running the factor analysis on selected data the value of fleet utilization is taking higher load of other factors. Thus, it has lower value than 0.4 (standard measure for detecting communalities in factor analysis.)All other factors are having higher computation of loading factors.

Table 1.4 Total Variance Explained

Comp	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Var.	CF %	Total	% of Var.	CF %	Total	% of Var.	CF %
1	4.312	33.168	33.168	4.312	33.168	33.168	3.749	28.842	28.842
2	3.119	23.992	57.160	3.119	23.992	57.160	3.379	25.994	54.836
3	1.598	12.295	69.455	1.598	12.295	69.455	1.900	14.619	69.455
4	.977	7.516	76.971						
5	.720	5.537	82.508						
6	.573	4.408	86.916						
7	.484	3.726	90.642						
8	.414	3.187	93.828						
9	.264	2.033	95.861						
10	.215	1.653	97.514						
11	.148	1.140	98.654						
12	.112	.863	99.517						
13	.063	.483	100						

Initial Eigenvalues: Eigenvalues are presenting the variances of the factors or parameters. The factor analysis is composed based on the correlation matrix, thus the variables are consistent, which means that the each variable has a variance of 1, and the total variance is equal to the number of variables used in the analysis, in last case 13.

Total – This segment contains the eigenvalues. The main factor will dependably represent the most difference (and subsequently have the most noteworthy eigenvalue), and the following component will represent as a significant part of the left finished change as it can, et cetera. Thus, each progressive factor will represent less and less difference.

% of Var. – This section contains the percent of aggregate difference represented by each factor.

CF% – This segment contains the combined level of variance represented by the present and every single going before factor. For instance, the third column demonstrates an estimation of 69.455. This implies the initial three factors together record for 69.455% of the aggregate change.

Extraction Sums of Squared Loadings – The quantity of lines in this board of the table relate to the quantity of variables held. In this case, we asked for that three variables be held, so there are three lines, one for each held factor. The qualities in this board of the table are ascertained similarly as the qualities in the left board, with the exception of that here the qualities depend on the normal difference. The qualities in this board of the table will dependably be lower than the qualities in the left board of the table, since they depend on the normal difference, which is constantly littler than the aggregate change.

Rotation Sums of Squared Loadings– The qualities in this board of the table speak to the conveyance of the change after the varimax turn. Varimax turn tries to amplify the difference of every one of the elements, so the aggregate sum of change represented is redistributed over the three separated elements.

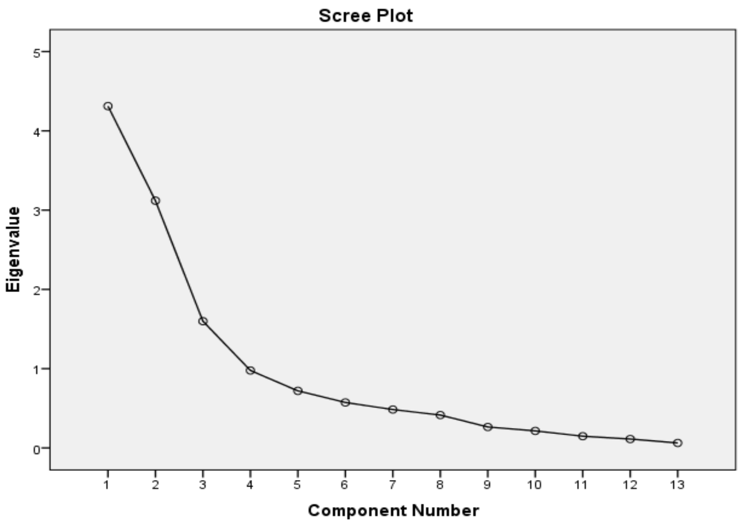

Figure 1.1 Scree Plot of Factor Loading

The scree plot diagram showsthe eigenvalue against the factor number. It can see these values in the initial two sections of the figure instantly above. From the third factor on, it can found that the line is level, which means the each progressive factor is representing littler and littler measures of the aggregate fluctuation.

6.2.4 Component Matrix

Interpretation for factor loadings is similar to interpretation of coefficients for principal component analysis. It is criterion, which can helps to determine which of these are large and which of these are considered negligible. This again is something of an arbitrary choice. The values that have considered large in boldface are using about 0.5 as the cutoff. Based on this determination will make the following statements:

Table 1.5 Component Matrix

	Component		
	1	2	3
Effective KM	-.365	**.837**	.122
No. of Passengers	-.571	**.714**	.171
Total EPKM	.565	.001	**.689**
Total CPKM	.325	-.214	**.741**
Loss	-.362	.439	**.515**
Load Factor %	.424	**.571**	-.118
Vehicle Utilized per day	**.688**	-.006	-.243
Fleet Utilization	-.033	.096	-.324
Crew Utilization	**.895**	.152	-.003
Diesel KMPL	**.649**	.634	.093
Engine Oil KMPL	.368	**.779**	-.237
Break Down	-.813	-.295	.139
Accidents	-.790	.464	-.062

The component matrix values can be interpreting as multiple R^2 values for regression models predicting the variables of interest from the 13 factors. The communality for a given variable can be interpreted as the proportion of variation in that variable explained by the 13 factors. In other words, if we perform multiple regression of GSRTC parameters against the 13 common factors, we obtain an R^2 = 89.5% for crew utilization, indicating that about 89.5% of the variation is explained by the factor model. The second leading parameter is Vehicle Utilized per day it explained 68.8% of the variation in factor one and similarly, the third leading parameter is Diesel KMPL it has explained 64.9% variation in factor one.

One assessment of how well this model is doing can be obtained from the communalities. This would indicate that the model explains most of the variation for those variables. In this case, the model does better for some variables than it does for others. The model explains explaining only about half of the variations for coefficient value lower than 0.5. The components having negative signs shows the decreasing value according to the selection of factor. An interpretation for each of the group variable for GSRTC parameters.

The second component explain that Effective KM having highest explained variation of 83.7%, total No. of Passengers explained 71.4% variation, Load Factor explains 57.1% and Engine Oil KMPL explained 77.9% variation for factor two. The third component covered by Total EPKM with 68.9%, total CPKM with 74.1% and Loss by 51.5% variation for factor three. This are the leading elements which may cause to disturb the GSRTC management for making batter operations.

6.2.5 Rotated Component Matrix

The rotated component matrix contains the pivoted factor loadings (factor design network), which speak to both how the factors are weighted for every performing component yet additionally the connection between the parameter and the factor.

Table 1.6 Rotated Component Matrix

	Component		
	1	2	3
Effective KM		**.872**	
No. of Passengers		**.920**	
Total EPKM			**.820**
Total CPKM			**.830**
Loss		**.706**	
Load Factor %	**.706**		
Vehicle Utilized per day			
Fleet Utilization			
Crew Utilization	**.764**		
Diesel KMPL	**.887**		
Engine Oil KMPL	**.816**		
Break Down			
Accidents		**.791**	

Since these are relationships, conceivable values run from -1 to +1. On the/organize subcommand, utilized the choice blank(.60), which advises SPSS not to print any of the connections that are .6 or less. This makes the yield less demanding to peruse by evacuating the messiness of low connections that are most likely not significant in any case.

7. Conclusion

The Rotated Component Matrix finally determines 10 listed parameters, which affects the operational aspects of GSRTC. The usage of Diesel KMPL highly associated with operational aspects. It shows 88.7% relation for first factor. Next to that Engine oil, KMPL relates 81.6% directly. Both the parameters associated with prices of petroleum products not under control by the GSRTC. The Crew utilization has direct impact of 76.4% and load factor of passengers is affects 70.6% to the top-level assessment of GSRTC. This are the internal factors among them two are beyond the control system of GSRTC. While two are manageable.

The second component studied that total number of passengers have highest affect the management i.e. 92%, of course it is difficult to manage such mass for sixteen depots of Gujarat state for administration of GSRTC. Moreover, another factor affects is effective kilometers operated by the GSRTC to provide services. It has direct impact of 87.2%, which is also difficult to manage. Higher density of traffic and load of passengers caused to increased number of accidents. It relates 79.1% to operational management and decision making to GSRTC. A margin or income affects much to the management, since long GSRTC running their services with biggest loss. This has direct association of 70.6% to the GSRTC.

The third components explain the factors, which are associated to operational management. It is total earning per kilometer it has effect of 82% and total cost per kilometer which has direct effect of 83%. Both the parameters explain the income and expenditure status of GSRTC. The income layer is little lower than the expenditure. It clearly shows that GSRTC already running their buses with 1%. This difference of loss in not covered other expenses of GSRTC to manage. Thus, it make clear status that the board of management has to reform their strategies which can provide proper services as well manages their own operations without loss.

References

1. Berenson M.L. and Levine D.M. (1996) *Basic Business Statistics,* Prentice-Hall, Englewood Cliffs, New Jersey.
2. Bhattacharyya, G. K., and R. A. Johnson, (1997). *Statistical Concepts and Methods*, John Wiley and Sons, New York.
3. Birnbaum, Z. W. (1952). "*Numerical tabulation of the distribution of Kolmogorov's statistic for finite sample size*", Journal of the American Statistical Association, 47, page 425.
4. Brown, L. D. Cai, T. T. and DasGupta, A. (2001). *Interval estimation for a binomial proportion*", Statistical Science, 16(2), 101-133.
5. Diamond, W. J. (1989). *Practical Experiment Designs,* Van-Nostrand Reinhold, New York.
6. Dixon, W. J. and Massey, F.J. (1969). *Introduction to Statistical Analysis*, McGraw-Hill, New York.
7. Draper, N. and Smith, H., (1981). *Applied Regression Analysis*, John Wiley & Sons, New York.
8. Fleiss, J. L., Levin, B. and Paik, M. C. (2003). *Statistical Methods for Rates and Proportions*, Third Edition, John Wiley & Sons, New York.
9. Guenther, W. C. (1977). *"Sampling Inspection in Statistical Quality Control",*, Griffin's Statistical Monographs, Number 37, London.
10. Hahn, G. J. and Meeker, W. Q. (1991). *Statistical Intervals: A Guide for Practitioners*, John Wiley & Sons, New York.
11. Hicks, C. R. (1973). *Fundamental Concepts in the Design of Experiments*, Holt, Rinehart and Winston, New York.
12. Hollander, M. and Wolfe, D. A. (1973). *Nonparametric Statistical Methods*, John Wiley & Sons, New York.
13. Howe, W. G. (1969). *"Two-sided Tolerance Limits for Normal Populations - Some Improvements",* Journal of the American Statistical Association, 64 , pages 610-620.

