Report On

A Statistical Analysis Of Road Accidents In India

Shubham Kumar

Bachelor of Science, Department of Statistics, AIASK, Amity University, Kolkata

Email ID: mail2shubham99@gmail.com

Abstract:

In the present days as well as the past, road accidents have always been of much social and economic concern. This project will be an attempt to investigate the causes behind road accidents in India and to compare fatalities, deaths, injuries, etc. The other aspect of the project includes the generation of a well suited model to check the effect of meaningful factors over the concerned dependent variable, inferring and any trend or seasonal pattern present in the time series and to finally fit a forecasting model for future purpose. The data used in the analysis has been collected from the official government website "data.gov.in" for some past years. Particularly, this fatal loss growing day by day in the social and economic aspect of our society is what which have motivated me to carry on this analysis.

We will fit some well suited model, (in this case ARIMA model) in accordance with the goodness of fit measure. We would also calculate the accident death rates (with respect to total population), accident death rate (with respect to total number of accidents), etc, fit appropriate Poisson models and to create an accident severity index over past years for which the data is available.

Hence, this statistical analysis is concerned mainly in arriving to suitable measures to effectively decrease the accident rates in the future and to forecast future accident numbers in India.

This project has been a great experience for me. There were various information's which I have gathered and has given me a broader picture to this field. This experience and exposure has helped my personal development. This experience has shown me a glimpse of how life is and will be in days to come.

Keywords: Poisson distribution, ARIMA model, VAR model, Multiple linear regression, ANOVA, goodness of fit (R-squared) measure, ADF test, Ljung-Box test, Z-test, AIC, ACF, PACF, Accident severity index.

Introduction:

Origin of the report:

This report, based on a two months study, helped me to gather practical information, which is necessary for my future life. I would like to express my deep respect to my institute guides of AIASK for giving me their valuable time and all the necessary guidance, which helped me to prepare this report.

Aims:

1. To look upon the descriptive charts and to make a comparison about road accidents in different states/UTs of India.
2. Poisson model fitting to accident death rates and accident injury rates over years.
3. Forecasting of number of road accidents in the future by the help of time series data of previous years.
4. Cross-sectional analysis at a single time point (2016) for prediction of total accident numbers with the help of significant causal variables.

Collection of Data:

All the datasets are collected from the excellent digital initiative of Indian government, an official govt. website "data.gov.in".

Organization:

The project has been divided into four sections and is organized as follows. Section 1 is a descriptive analysis about the different charts, rates and severity index. Section 2 is an attempt to fit a appropriate probability distribution (Poisson) to accident death rates and accident injury rates. Section 3 elaborates about the methodologies and diagnostic procedures to about the best fitted ARIMA(p,d,q) model according to the given time series data. Furthermore, a VAR modeling is also applied the data being multivariate time series. And, finally section 4 describes about a cross-sectional analysis for a single time point of collected data.

All the analysis and computations has been carried out in R (a statistical analytical software), with the help of the following packages:

a. tseries
b. forecast
c. lmtest
d. vars

Section 1

In the first section, we have done an extensive comparison of number of accidents according to different states and union territories of India for a particular year (2016) and have seen that Tamil Nadu has the highest percentage of road accidents (14%) followed by Madhya Pradesh (11%), Karnataka(9%), etc.

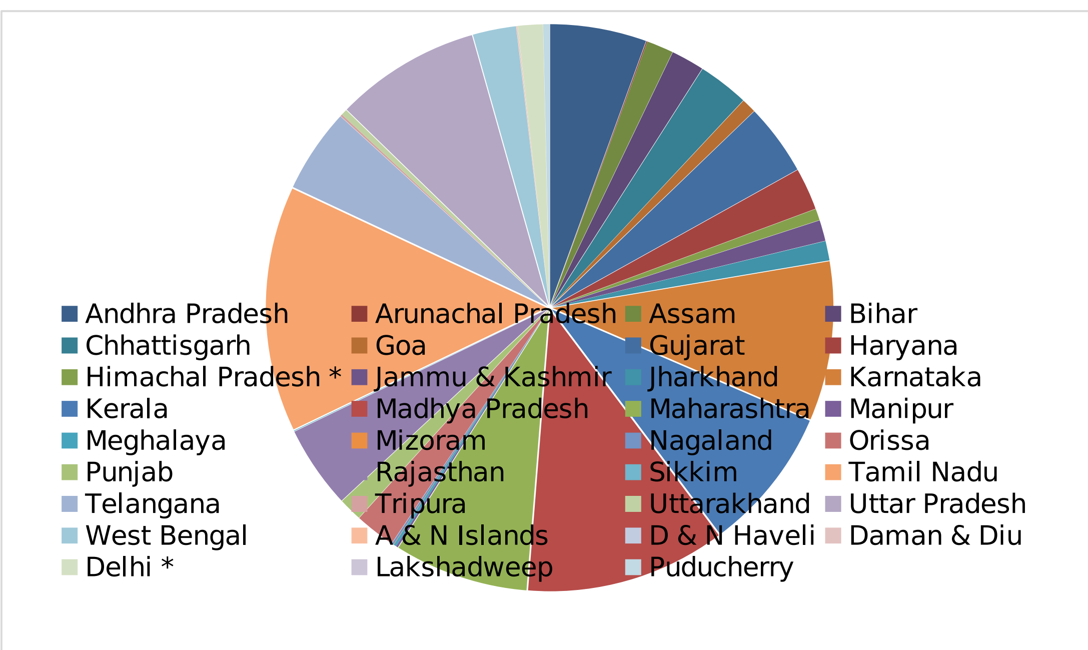


Figure 1: Pie Chart of no. of road accidents in different states & territories (2016)

From the plots mentioned below, we can get a basic idea about the type of accidents and fatality level of the accidents according to the different states. We can notice that minor injury accidents are most relevant among the dataset and grievous injury accidents can be considered to be second most relevant. So, we can interpret that minor injuries and most common in road accidents and the other most common action being grievous injuries.

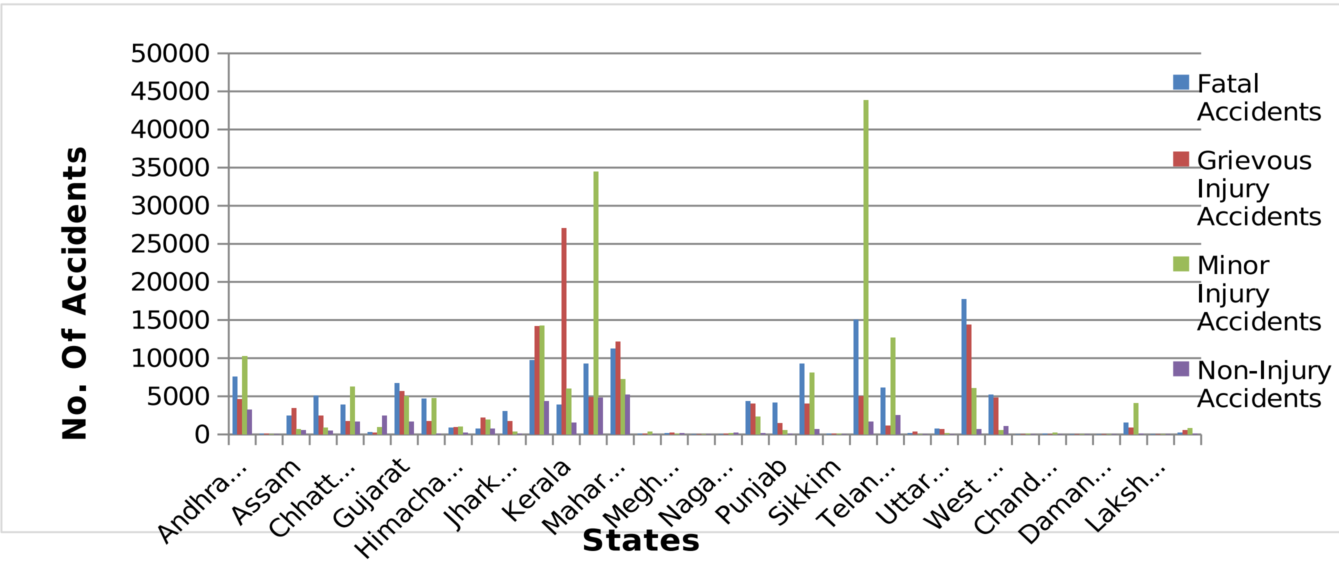
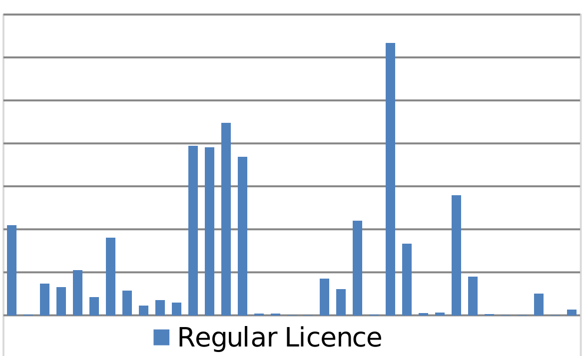


Figure 2: Types of accidents on the rate of fatality of the different states and territories (2016)

One more interpretation we can incur from the plots is about the no. of accidents caused by legal license holders, learner's license holders and people driving without any license. As we can notice, the most number of accidents caused by legal license holder as well as driving without license has happened at Tamil Nadu and most number of accidents due to learner's license holders has happened at Madhya Pradesh.



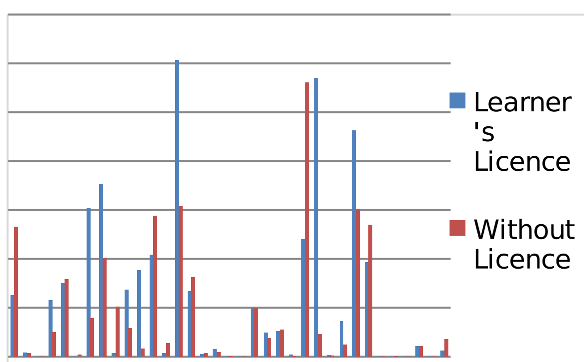


Figure 3: Accidents due different type of license holders

Death rate, injury rate and accident severity index:

Accident death rate percentage with respect to total population (1994-2016):

The accident death rate (with respect to total population) is been computed by the formula:

$ADR_1 = 100)$ %

Accident death rate percentage with respect to total number accidents in the year (1994-2016):

$$ADR_2 = \left(\frac{Total\,no.\,of\,accident\,deaths \in the\,year}{Total\,number\,of\,accidents\,happened \in the\,year} \times 100\right)\%$$

Accident injury rate percentage with respect to total population (1994-2016):

$$AIR = \left(\frac{Total\,no.\,of\,accident\,injuries}{Total\,p\,opulation} \times 100\right)\%$$

Table 1: Accident death rates & Injury rates

ADR_1 (%)	ADR_2 (%)	AIR (%)
0.007131	19.78218	0.034458
0.007657	20.1083	0.034965
0.00793	20.11428	0.039243
0.00802	20.60021	0.039421
0.008171	20.75721	0.039943
0.008228	21.20966	0.037651
0.007776	20.15869	0.039343
0.007864	19.94098	0.039395
0.008099	20.77905	0.039091
0.008095	21.14396	0.040957
0.008583	21.54358	0.043046
0.008667	21.62024	0.042464
0.009508	22.94303	0.04464
0.010141	23.88151	0.045488
0.010471	24.72849	0.045704
0.010825	25.83555	0.044405
0.011431	26.92263	0.044828
0.011774	28.6295	0.042257
0.011444	28.19388	0.042187
0.011243	28.2793	0.040446
0.011274	28.53923	0.039832
0.011653	29.14366	0.039894
0.011883	31.37093	0.038979

Accident severity index:

The accident rates are plotted over years and the rapid growth of road accidents in India has been noticed over from 1994 to 2016. Accident severity index measures the seriousness of accidents in the country.

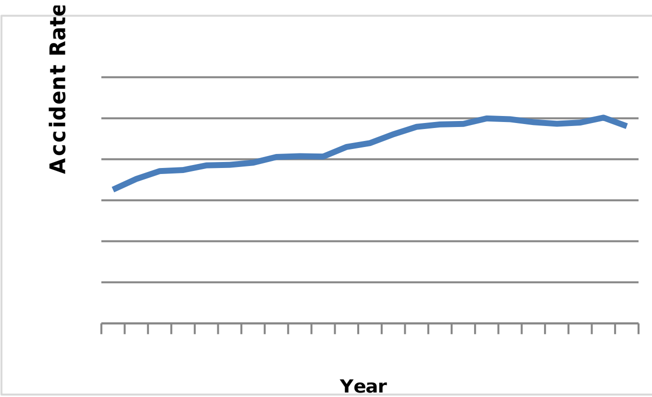


Figure 4: Accident Severity Index

Section 2

In this section we will fit appropriate Poisson distributions to ADR_1, ADR_2, AIR (computed in the previous section).

Poisson distribution: A Poisson distribution is a discrete modeling the number of events occurring at a fixed interval of time or space.

Mathematically, if X P (λ) [i.e. X is a discrete random variable following Poisson distribution with parameter λ), then its probability mass function (PMF) is given by,

P(X=x) = $\frac{e^{-\lambda} \times \lambda^{x}}{x!}$, for all values discrete positive values of x.

Properties of Poisson distribution:

Mean of the distribution, E(X) = λ

Variance of the distribution, V(X) = λ

Model fitting for ADR_1, ADR_2, AIR:

Assumption: Though the concerned variables are continuous, we would treat team as isolated values as the approximation by the Poisson fitting is quite accurate and due to the sake of simplicity of calculations.

We have simply taken the averages of ADR_1, ADR_2 & AIR and treated them as parameters to fit the corresponding Poisson models.

Average(ADR_1)= 0.00947249, Average(ADR_2)=23.7489585, Average(AIR)= 0.040810271

Define, ADR_1 = X, ADR_2=Y, AIR=Z (say)

Hence, X P(0.00947249), Y P(23.7489585), Z P(0.040810271)

One can compute the desired probabilities for given values of x, y or z.

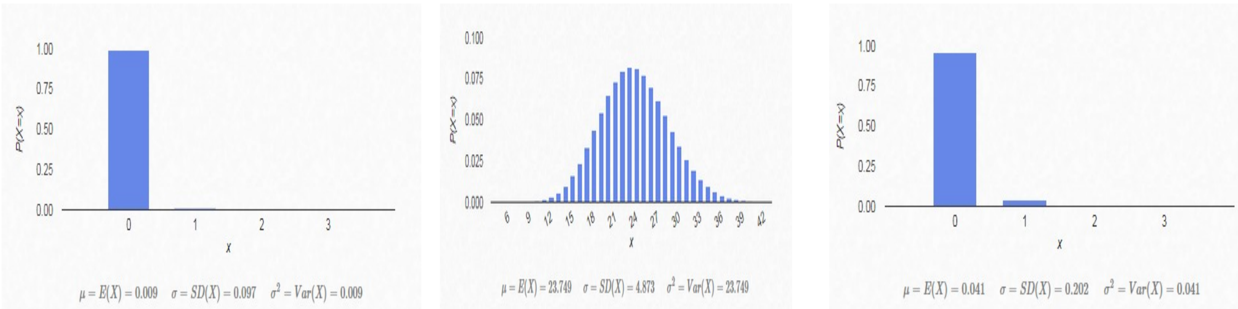


Figure 5: Density Plots of the fitted Poisson distributions

Section 3

Objective: In this section, our objective is to forecast the number of road accidents in the future by the help of time series data of previous years.

Data & Methodology:

The analysis involves yearly data on the road accidents in India for 23 consecutive years (1994-2016). Based on this data, we have tried we have fitted a suitable stochastic models, ARIMA and VAR model by assuming the data to be univariate and multivariate in two different cases respectively.

Univariate time series analysis: A univariate time series refers to a time series that consists of single observations recorded over regular time intervals. In univariate time series analysis, we consider only one variable and try to predict it by modeling its dependence upon its own values for different previous equidistant time points. Eg: Monthly return data of a stock in stock market.

At first, we are considering our dataset of number of road accidents to be univarite and will try to forecast the future values. For this case, an Autoregressive Integrated Moving average (ARIMA(p,d,q)) model will be fitted.

Now, we need to compute correlograms of autocorrelation function (ACF) and partial autocorrelation function (PACF) between the members of the series.

Autocorrelation: In statistics, the autocorrelation of a real or complex random process is the Pearson correlation between values of the process at different times, as a function of the two times or of the time lag.

Partial Autocorrelation: In time series analysis, the partial autocorrelation function (PACF) gives the partial correlation of a stationary time series with its own lagged values, regressed the values of the time series at all shorter lags. It contrasts with the autocorrelation function, which does not control for other lags, i.e. PACF is some kind of ACF but eliminating effects of other lags.

ARIMA Modelling:

ARIMA models are the most general class of models for forecasting a time series which can be residualised by transformations such as differencing, etc. ARIMA stands for “Auto-Regressive Integrated Moving Average”. A non-seasonal ARIMA model is classified as ARIMA(p,d,q) model, where,

a. p is the parameter of autoregressive model (i.e. number of autoregressive terms)
b. d is the number of differences.
c. q is the parameter of moving average model (i.e. number of lagged forecast errors in the prediction).

In general, an ARIMA(p,d,q) model is defined as,

$$X_t - \alpha_1 X_{t-1} - \cdots - \alpha_{p'} X_{t-p'} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

Box-Jenkins Method:

We will use Box-Jenkins procedure to fit the well suited ARIMA model for our data which consists of:-

1. Identification
2. Estimation
3. Diagnostic checking

Identification: In this step, we identify about the stationarity of the data and infer about the values of p & q of the model by comparing the correlograms of the time series data with the theoretical ACFs and PACFs. Hence, in general the ACF & PACF tells us that how many lags are appropriate for our model. The general characteristics of the theoretical ACFs and PACFs are as follows:-

Table 2: Characteristics of ACFs & PACFs

Model	ACF	PACF
AR(p)	Spikes and decays down	Spikes and cut off after lag p

MA(q)	Spikes and cut off after lag q	Spikes and decays down
ARMA(p,q)	Spikes and decays down	Spikes and decays down

Estimation: Several methods are available for estimating the parameters of the ARMA model depending on the assumptions of the error terms. Some of the reputed methods are:

a. Yule Walker procedure
b. Method of Moments
c. Maximum likelihood estimation, etc.

The estimations, being tedious are usually carried out by computer softwares or programs.

Diagnostic Checking: The best model is obtained by following various tests, such as,

a. The model with the lowest value of AIC(Akaike information criterion)/BIC(Bayesian Information criterion)/SBIC(Scharz Criterion) is chosen as the best model.
b. Plotting the model residuals and checking for randomness by the help of Ljung-Box test.
c. Coefficient significance test of the model.
d. Checking the accuracy measures of the model.

Results:

Identification:

We proceed with our objective of analyzing the time series in the following steps. Firstly, we will check the stationarity of our data. By plotting the time series data the presence of trend can simply be identified.

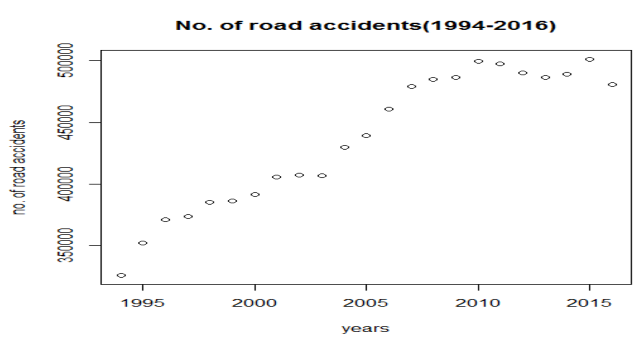


Figure 6: Time series plot of no. of accidents

Then also, to be sure we will apply Augmented Dickey Fuller Test, where our null hypothesis is the data to be explosive vs. alternative hypothesis is the data to be stationary. And, by running the test in R, we got the p-value for the test to be

0.9686 (greater than 0.05) and hence we have accepted the null hypothesis (i.e. the data is explosive). So, we have done differencing in the series to make it stationary. According to the AIC (Akaike information criteria) we have seen that I(2) [differencing two times} has lower measure of AIC and hence will use ARIMA model with d=2 (p-value of ACF test=0.03).

Now, we plot the ACF and PACF of the data and by observing the correlograms can easily infer about the p & q values. For the data, p=0 and q=1.

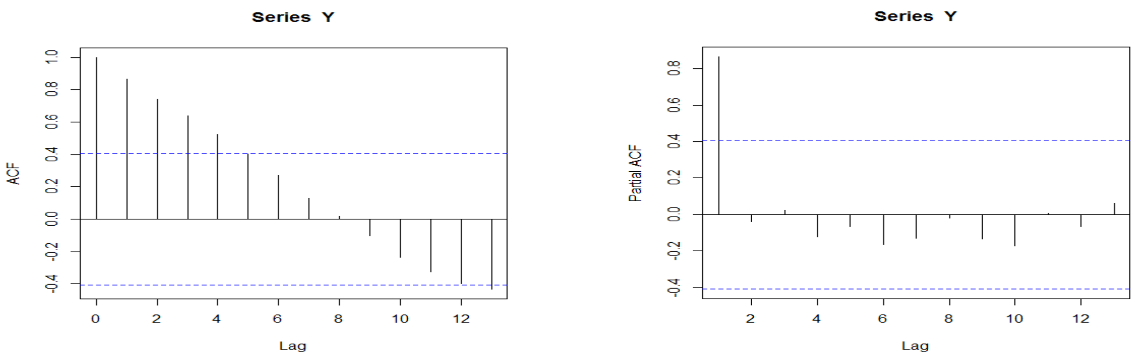


Figure 7: ACF & PACF of concerned time series

Hence, ARIMA(0,2,1) is obtained as our required model which we will be using as a estimate to predict the future values of number of road accidents.

The generalized ARIMA(0,2,1) model is:

$Y_t = 2Y_{t-1} - Y_{t-2} - a(e_{t-1})$

```
ARIMA(0,2,1)

Coefficients:
          ma1
      -0.6819
s.e.   0.2097

sigma^2 estimated as 128421087:  log likelihood=-225.64
AIC=455.28   AICc=455.95   BIC=457.37
```

Figure 8: Output of Best fitted ARIMA model

Thus, from the above output, we can express our model as,

$Y_t = 2Y_{t-1} - Y_{t-2} + 0.6819(e_{t-1})$

Actually, this is nothing but linear exponential smoothing model.

Estimation:

The estimation of the parameter has been carried by the help of R software.

Diagnostic Checking:

The model has the lowest AIC (=455.28) among all other significant fitted ARIMA models. The value of the accuracy measures are also best in the model as compared to other models.

```
       ME      RMSE       MAE         MPE      MAPE       MASE
-3451.662 10567.43 8577.536 -0.8134975 1.938454 0.8417752
       ACF1
-0.06943912
```

Figure 9: Accuracy measures

Z-test for significance of coefficient:

```
z test of coefficients:

    Estimate Std. Error z value Pr(>|z|)
ma1 -0.68192    0.20971 -3.2517 0.001147 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 10: Output of Coefficient significance test

From the above output, we can see that the coefficient of ma_1 is significant at 5% level of significance as the p-value=0.001(<0.05). Hence, we reject the coefficient to be zero and consider the computed value.

Ljung-Box Test:

Here, H_0: Residuals of the model are i.i.d. normally distributed with zero mean and constant variance.

Vs. H_1: not H_0

```
        Box-Ljung test

data:  resid_x
X-squared = 4.1615, df = 10, p-value = 0.9398
```

Figure 11: Output of Ljung-Box test

So, we can see that the p-value of the test=0.9390(>0.05) [at 5% level of significance] we accept the null hypothesis, i.e. the residuals are i.i.d normally distributed with 0 mean and constant variance.

Consequently, we can conclude that our model is the best fit and can forecast the future values with the help of it.

Forecast:

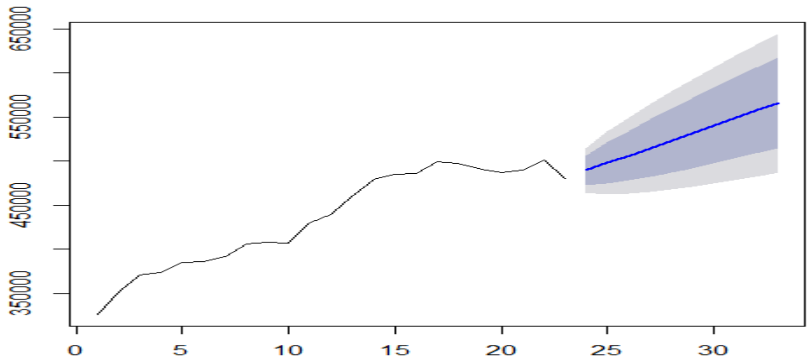


Figure 12: Forecasted values for the next 10 years

So, we can notice from the above plot that our model is quite efficient and the forecasting is nicely done.

Now, as there are other variables also present in the data which may have effect on our desired variable our data can also considered to be a multivariate time series model.

Multivarite time series analysis: When we compose univariate time series model along with structural models then it is known as a multivariate time series model.

For this case we will fit a Vector Autoregressive model to our data, and will try to find out the dependency of No. of road accidents on its own previous lags along with influence of Population in India in that year, length of road (kms), no. of registered motor vehicles of that year.

Hence, we have fitted a best suited VAR model by using optimal lag selection technique and created a very well fitted value with goodness of fit value (R-squared) of 0.9794 which means a very good fit.

Fitted model:

Y = Y.l1 + Population.of.India..in.thousands..l1 + Total.Number.of.Registered.Motor.Vehicles..in.thousands..l1 + Road.Length..in.kms..l1 + Y.l2 + Population.of.India..in.thousands..l2 + Total.Number.of.Registered.Motor.Vehicles..in.thousands..l2 + Road.Length..in.kms..l2 + constant

Where, Y=Number of road accidents in a year

Estimates and p-values of the independent variables:

```
                                                        Estimate
Y.l1                                                   8.596e-01
Population.of.India..in.thousands..l1                  2.747e-01
Total.Number.of.Registered.Motor.Vehicles..in.thousands..l1 -1.479e+00
Road.Length..in.kms..l1                                6.577e-02
Y.l2                                                  -3.468e-01
Population.of.India..in.thousands..l2                  1.416e-01
Total.Number.of.Registered.Motor.Vehicles..in.thousands..l2  8.466e-01
Road.Length..in.kms..l2                               -4.768e-02
const                                                 -2.466e+05
```

Figure 13: Estimates of coefficients

```
                                                             Pr(>|t|)
Y.l1                                                          0.00497 **
Population.of.India..in.thousands..l1                         0.49106
Total.Number.of.Registered.Motor.Vehicles..in.thousands..l1   0.09722 .
Road.Length..in.kms..l1                                       0.07472 .
Y.l2                                                          0.31536
Population.of.India..in.thousands..l2                         0.71294
Total.Number.of.Registered.Motor.Vehicles..in.thousands..l2   0.39216
Road.Length..in.kms..l2                                       0.13661
const                                                         0.21221
```

Figure 14: P-values for testing significance of the coefficients

Hence, from the above p-value chart we can conclude about the significance of the independent variables and can include the significant variables accordingly in the model.

Goodness of Fit measure:

```
Residual standard error: 8875 on 12 degrees of freedom
Multiple R-Squared: 0.9794,     Adjusted R-squared: 0.9656
```

Figure 15: output of r-squared measures

Hence, we have created a well fitted forecasting model that can be used for predicting future values.

Section 4

Objective: In this section we will carry out a cross-sectional analysis at a single time point (2016) for prediction of total accident numbers with the help of significant causal variables.

Data & Methodology:

The data has been collected from the website of “data.gov.in” and consists of Total number of road accidents, total number license issued, total number of accidents caused without license for the different states and UTs of India on 2016.

We will run a multiple linear regression analysis for predicting total number of accidents while taking explanatory variables to be total number license issued, and total number of accidents caused without license. Firstly, we have checked that both chosen explanatory variables has significant effects on the dependent variable by using Chi-square test of independence.

Structure of the model:

We will fit a multiple linear regression of the form,

$X_1=a_0+a_1X_2+a_2X_3+e$, [where, errors (e)s are i.i.d. normally distributed random variables with 0 mean and constant variance]

Results & Interpretation:

Define,

X_1: Total number of accidents (dependent variable)

X_2: Total number of license issued (independent variable)

X_3: Total number of accidents caused without license (independent variable)

```
Call:
lm(formula = X1 ~ X2 + X3)

Residuals:
   Min     1Q Median     3Q    Max 
-15296  -2480  -1229   1989  26692 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1.133e+03  1.866e+03   0.607  0.54859    
X2          8.878e-03  2.436e-03   3.645  0.00108 ** 
X3          9.540e+00  1.221e+00   7.814 1.64e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7861 on 28 degrees of freedom
Multiple R-squared:  0.8189,    Adjusted R-squared:  0.806 
F-statistic: 63.32 on 2 and 28 DF,  p-value: 4.073e-11
```

Figure 16: Outputs Of multiple linear regression

Our estimated regression equation is: $X_1 = (1.13\times10^{3}) + (8.88\times10^{-3})X_2 + 9.54X_3$

From the above output, we can see, if we want to test $H_0{:}a_0=0$ vs. $H_1{:}a_0\neq 0$

We can see, at 5% level of significance, we will have to accept the null hypothesis (p-value=0.548) and hence we can say that the constant term of the model is not very significant.

But, as one can notice from the above output, both of the other regressors are significant at 5% level.

Hence, we can observe from the fitted regression model that total accident number of a year is highly influenced by the no. of accidents caused without license and has a strong positive correlation which is quite logical. Total number of registered vehicles also has a positive effect on the total number of road accidents in that year.

One can use this model to predict the possible number of accidents due to certain values of total number of registered vehicles on a year and number of accidents caused by drivers without license. Furthermore, our government can take appropriate measures and checks with the help of this model.

Conclusion:

With the help of the analysis, different forecasting models (ARIMA & VAR) for predicting the future road accidents have been created which can be extensively used by government to take appropriate measures accordingly. There is also a descriptive analysis included in the paper to give an individual a rough idea about the current scenario of road accidents in India.

Some well assumed Poisson models are fitted to the death rates and injury rates as well for computing concerned probabilities. At the end, there is a descriptive analysis for the prediction of total road accidents at a certain time point due to some discussed causal variables.

In short, this project is a whole hearted attempt to fight back against this fatal socio-economic loss growing day by day in our society.

Acknowledgement:

I have got the opportunity to work on discussed topic. So, I acknowledge the valuable contribution of my internal guides and other respected faculties

without whose guidance, support and co-operation my project would not have been possible.

My sincere gratitude towards my family members who have supported me in all aspects of my life as well as this project. This project has been a great learning experience for me, which helped me enhance my skills, strength and confidence for future opportunities.

References:

1. IRJET: Journal of Statistical data analysis of road accidents - Mukul Nama, Nandeshwar Lata,

Dr. Bharat Nagar

2. Road accidents journal in India (2018) - Govt. Of India

3. Forecasting of demand using ARIMA model - Jamal Fattah, Latifa Ezzine, Zineb Aman

4. National statistics of road accidents in India - Manisha Ruikar

5. The analysis of time series: Theory and Practice - C. Chatfield

Appendix:

Table 3: Yearly data of road accidents in India (1970-2017):

Years	Total Number of Road Accidents (in numbers)	Total Number of Persons Killed (in numbers)	Total Number of Persons Injured (in numbers)	Population of India (in thousands)	Total Number of Registered Motor Vehicles (in thousands)	Road Length (in kms)
1970	114100	14500	70100	539000	1401	1188728
1980	153200	24000	109100	673000	4521	1491873
1990	282600	54100	244100	835000	19152	1983867
1994	325864	64463	311500	904000	27660	2890950
1995	351999	70781	323200	924359	30295	2975035
1996	371204	74665	369502	941579	33786	3202515
1997	373671	76977	378361	959792	37332	3298788

1998	385018	79919	390674	978081	41368	3228356
1999	386456	81966	375051	996130	44875	3296650
2000	391449	78911	399265	1014825	48857	3316078
2001	405637	80888	405216	1028610	54991	3373520
2002	407497	84674	408711	1045547	58924	3426603
2003	406726	85998	435122	1062388	67007	3528654
2004	429910	92618	464521	1079117	72718	3621507
2005	439255	94968	465282	1095722	81502	3809156
2006	460920	105749	496481	1112186	89618	3880651
2007	479216	114444	513340	1128521	96707	4016401
2008	484704	119860	523193	1144734	105353	4109592
2009	486384	125660	515458	1160813	114951	4471510
2010	499628	134513	527512	1176742	127746	4582439
2011	497686	142485	511394	1210193	141865.6	4676838
2012	490383	138258	509667	1208116	159490.6	4865394
2013	486476	137572	494893	1223581	181508	5231922
2014	489400	139671	493474	1238887	190704	5402486
2015	501423	146133	500279	1254019	210023	5472144
2016	480652	150785	494624	1268961	230031	5603293
2017	464910	147913	470975	1283601	NA	NA

S. No.	State/ UT	Regular Licence	Learner's Licence	Without Licence	Total numbers of license issued (2016)	Total Accidents
1	Andhra Pradesh	20973	1254	2661	398773	25727
2	Arunachal Pradesh	93	85	71	24488	241
3	Assam	7435	0	0	116577	7170
4	Bihar	6559	1160	503	486469	8855
5	Chhattisgarh	10485	1507	1588	336469	13563
6	Goa	4263	1	40	37438	3917
7	Gujarat	18026	3042	791	2894328	19081
8	Haryana	5686	3530	2018	356295	11258
9	Himachal Pradesh *	2301	68	1026	80424	3114
10	Jammu & Kashmir	3544	1369	588	65161	5624
11	Jharkhand	3002	1768	162	194957	5198

Table 4: Count of accidents due to type of license holdersand total number of license issued at different states/UTs (2016)

12	Karnataka	39436	2084	2883	1274081	42542
13	Kerala	39072	76	272	906684	38470
14	Madhya Pradesh	44822	6068	3082	890237	53399
15	Maharashtra	36905	1342	1631	1874915	35853
16	Manipur	414	53	71	70715	578
17	Meghalaya	375	154	91	45047	675
18	Mizoram	73	0	10	NA	68
19	Nagaland	75	0	0	19111	531
20	Orissa	8529	1005	998	371724	10855
21	Punjab	6073	498	381	NA	6273
22	Rajasthan	21991	521	554	878792	22112
23	Sikkim	157	45	8	7053	196
24	Tamil Nadu	63421	2400	5610	1013636	65562
25	Telangana	16650	5702	459	721267	22484
26	Tripura	498	36	23	108060	503
27	Uttarakhand	613	728	250	145393	1603
28	Uttar Pradesh	27960	4627	3025	1451440	38783
29	West Bengal	8946	1937	2697	NA	11631
30	A & N Islands	238	0	0	11526	189
31	D & N Haveli	62	0	8	9756	67
32	Daman & Diu	64	0	7	5717	79
33	Delhi *	5058	221	217	583785	6673
34	Lakshadweep	1	0	0	1816	1
36	Puducherry	1279	124	363	NA	1693

Table 5: Types Of accidents and total accident count according to different states and UTs (2016)

States/UTs	Fatal Accidents	Grievous Injury Accidents	Minor Injury Accidents	Non-Injury Accidents	Total Accidents
Andhra Pradesh	7564	4607	10285	3271	25727
Arunachal Pradesh	103	86	26	26	241

Assam	2474	3451	706	539	7170
Bihar	5045	2431	887	492	8855
Chhattisgarh	3878	1706	6285	1694	13563
Goa	306	237	926	2448	3917
Gujarat	6739	5653	5033	1656	19081
Haryana	4700	1700	4771	87	11258
Himachal Pradesh	907	959	1042	206	3114
Jammu & Kashmir	765	2179	1945	735	5624
Jharkhand	3034	1734	337	93	5198
Karnataka	9739	14191	14247	4365	42542
Kerala	3915	27034	5994	1527	38470
Madhya Pradesh	9258	4863	34493	4785	53399
Maharashtra	11220	12164	7253	5216	35853
Manipur	107	91	342	38	578
Meghalaya	140	253	101	181	675
Mizoram	55	8	2	3	68
Nagaland	35	72	159	265	531
Orissa	4372	4021	2302	160	10855
Punjab	4139	1490	561	83	6273
Rajasthan	9300	4017	8110	685	22112
Sikkim	60	70	58	8	196
Tamil Nadu	15061	5005	43856	1640	65562
Telangana	6110	1165	12695	2514	22484
Tripura	153	339	2	9	503
Uttarakhand	727	674	167	35	1603
Uttar Pradesh	17706	14363	6044	670	38783
West Bengal	5199	4811	560	1061	11631
A & N Islands	20	54	88	27	189
Chandigarh	103	7	198	34	342

D & N Haveli	40	20	4	3	67
Daman & Diu	36	31	4	8	79
Delhi	1565	907	4110	91	6673
Lakshadweep	0	1	0	0	1
Puducherry	221	577	807	88	1693