ISSN: 2456–5474 RNI No.UPBIL/2016/68367 Vol.-6* Issue-2* March- 2021
Innovation The Research Concept

Analysis of Accidental Data since 2005 to 2017 regarding the Number of Accidents Deaths on National highway State Highway and other Roads in India

Paper Submission: 05/03/2021, Date of Acceptance: 15/03/2021, Date of Publication: 25/03/2021

Kunal Agrawal
Student
Berkarlay University
California. U.S.

P.C.Gupta
Retired Professor of Stat.
V.N.South Gujarat University
Surat, Gujrat, India

The data regarding the accidents and death there in is being analyse for the country as a whole .Exponential and Modified Exponential have been used to predictnumber of accidents deaths on National highway state highway and other roads in India.

Keywords: Number of Traffic Accidents, Deaths in accidents, National highway.

Introduction

The aim of this exploration is to use statistical methods to predict the number of accidents, road death, and injuries occurred for 2018 to 2021, using data from 2005 to 2017. The annual data provided by the government of India is available from 2005 to 2017. We will predict the number of fatalities, number of persons injured, and number of road accidents (as given in the source data) in national highways and state highways for the years 2018 to 2021.Based on the data, we will predict the number of accidents deaths on National Highways and State Highways:

In order to do the prediction, we must make some hypotheses about the source distribution of the accident data. Based on decades of studies, accident data, like demographic data, generally follows an exponential or modified exponential distribution.1

We will fit the two models 1. Exponential and 2. Modified Exponential to the source data; we will test the validity of the fit using the Chi-Squared test of goodness of fit.

Background Information

To conduct this exploration, we must first understand the theory behind the statistical predictions.

Exponential

Exponential function is modeled in the form-

$Y = a' e^{bt}$

(1)

Where Y is the variable to be estimated and t is the transformed independent variable (the time has been transformed from the origin year which is 2011 in this case). 2011 becomes year 0, and 2010 is t = -1, 2012 is t = 1 etc. a' and b are constants. t is transformed in order to simplify the calculation, as we will see later on.

I stands for i^{th} observation and it ranges from 1 to n, '^' means that the value is estimated. E.g. $\hat{Y}$stands for estimate of Y.

$\ln Y = \ln a' + bt$

(2)

$z_i = a + bt_i + e_i$

(3)

Where $z_i = \ln Y_i$ and a islna' and e_i is the error associated with the i^{th} observation.

By minimizing $\sum e_i^2$ with respect to choice of a and b, we minimize $\sum (z_i - a - bt_i)^2$ for the optimum choice of a and b. This is done by equating the first derivative of $\sum e_i^2$with respect to a and b to zero to give us the normal equation. The second derivative should be negative. We want to minimize $\sum e_i^2$ as e_i (error) are identically independently normally distributed with mean zero and constant variance σe^2.

Derivative:

$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(z_i - a - bt_i)^2$

(4)

*From now on, when I put Σ, I refer to $\sum_{i=1}^{n}$

$$\frac{d\sum e_i^2}{da} = 2\sum(z_i - a - bt_i)$$

Equating to zero for first derivative.

$$2\sum(z_i - a - bt_i) = 0$$

$$\sum z_i - na - b\sum t_i = 0$$

$\sum z_i = na + b\sum t_i$ (5)

Where n is the number of observations.

Now we derivate with respect to b-

$$\frac{d\sum e_i^2}{db} = 2\sum(z_i - a - bt_i)t_i$$

Equating it to zero for derivative:

$$2\sum(z_i - a - bt_i)t_i = 0$$

$$\sum z_i t_i - a\sum t_i - b\sum t_i^2 = 0$$

$\sum z_i t_i = a\sum t_i + b\sum t_i^2$

(6)

$\sum z_i = na + b\sum t_i$ and $\sum z_i t_i = a\sum t_i + b\sum t_i^2$ are the normal equations for the exponential model.

The solution to the normal equation will estimate the parameters of regression; in this case, it will give us the values of a and b.

$\sum z_i = na + b\sum t_i$

(7)

$\sum z_i t_i = a\sum t_i + b\sum t_i^2$ (8)

Which gives

$\hat{a} = \overline{z_n} - \hat{b}\bar{t}_n$

(9)

Where $\bar{z}_n = \sum\frac{z}{n}$ which is the sample mean of character z, and $\bar{t}_n = \sum\frac{t_i}{n}$ which is the sample mean of character t. Sample mean is defined as the

$\hat{b} = \frac{\sum z_i t_i - (\sum z_i)(\sum t_i)}{\sum t_i^2 - (\sum t_i)^2}$

(10)

And since t_i is a transformed variable to origin year 2011, the sum of t_i or $\sum t_i$ is actually equal to zero. Therefore, in this case, the equations above can be further simplified to:

$\hat{b} = \frac{\sum z_i t_i}{\sum t_i^2}$

(11)

$\hat{a} = \overline{z_n}$

(12)

Since a = lna', a' can be written as-

$$\widehat{a'} = e^{\hat{a}}$$

And b can be written as-

$\widehat{b'} = e^{\hat{b}}$

So, our prediction function is

$\hat{Y} = \widehat{a'}e^{\hat{b}t}$

(13)

The method consists of minimizing $\sum e_i^2$ to determine the values of a and b.

Modified Exponential

Modified Exponential is modeled in the form

$Y = a'(b')^t$

(14)

Where (Y, t) is the pair of observation of dependent variable (Y) and independent variable (t). t is the transformed time with origin year of 2011.

The equation above can be equated to-

$\ln Y = \ln a' + t\ln b'$

(15)

Which can also be written as?

$z = a + bt$ (16)

Where a= lna' , b = lnb' and z = lnY.

Let

$z_i = a + bt_i + e_i$

(17)

Where e_i is the error associated with the i^{th} observation (i = 1 to n). In the method of least square we minimize $\sum e_i^2$ for determination of a and b, which are obtained by the following equations. One crucial property of this method is that it provides unbiased minimum variance and unbiased estimators of a' and b' respectively. Variance is defined as the average of the squared differences from the mean.

The solutions of the equations provide us with an estimate of a & b.

$\sum z_i = na + b\sum t_i$

(18)

$\sum z_i t_i = n\sum t_i + b\sum t_i^2$

(19)

The rest follows same as in exponential method, but in this case $b' = e^b$.

$\hat{b} = \frac{\sum z_i t_i - (\sum z_i)(\sum t_i)}{\sum t_i^2 - (\sum t_i)^2}$

(20)

$\hat{a} = \bar{z}_n - \hat{b}\bar{t}_n$

(21)

Thus, our fitted model is

$\hat{Y} = \widehat{a'}(\widehat{b'})^t$

(22)

To determine whether the function which is used as a predictor of Y' for a given value of X' provides accurate values using Chi Square test of goodness of fit.

Chi Squared Test

Chi Squared test is a technique in statistical theory that enables us to examine the appropriateness or "goodness" of the function that has been fitted. It can also be used to test the independence of two attributes, or how much the occurrence of one variable effects the other variable.

The Chi Square test for goodness of fit is modelled by

$\sum\frac{(Y_i - \hat{Y}_i)^2}{\hat{Y}_i}$

(23)

$\hat{Y}_i$ Will the predicted variable, Y_i is the observed variable.

The chi square test will be used on every year from 2005 to 2017. In order to predict for the years 2018 to 2021, we create estimations for 2005 to 2017 by fitting the respective models. We compare the estimations ($\hat{Y}_i$) to the observations (Y_i) and do the chi square test.

The chi square distribution has n-p-1 df (degrees of freedom), where n is the number of observations, p is the number of parameters to be estimated. Degrees of freedom are the number of variables in the calculation that are free to vary. In this study, there are only two parameters, a and b, therefore p is 2, and we have 13 observations from 2005 to 2017, therefore n is 13. Hence the chi square distribution for this study has 10 df.The models have been designed to allow a 5 percent of error.

From the table in the annex, the tabulated value of chi square at 10df (degrees of freedom) and 5 percent level of significance is 18.31. This means that for our model to be considered a good fit, the calculated value of chi square must be less than or equal to 18.31.

Confidence Interval Estimation

A confidence interval is used to indicate the reliability of the estimate ofa population parameter.

Interval Estimation is the use of sample data to calculate an interval of possible values of an unknown population parameter.2

If we assume the data follows a normal distribution, which we can because the data is demographic data, then the we know that if x_1 x_2…x_n is a random sample from $N(\theta,\sigma^2)$, with θ being a parameter σ^2 being the variance of the values, we can say it follows

$$\frac{\sqrt{n}(\bar{x}-\theta)}{\sigma}$$

Which is the z-score and follows for N(0, 1). We can build the interval for which θ lies with

$$P\left[-2.576<\frac{\sqrt{n}(\bar{x}-\theta)}{\sigma}<2.576\right]=0.99$$

Or

$$P\left[\bar{x}-\frac{2.576\sigma}{\sqrt{n}}<\theta<\bar{x}+\frac{2.576\sigma}{\sqrt{n}}\right]=0.99$$

(24)

If the standard deviation of the data (sigma) is known, then $\bar{x}\pm\frac{2.576\sigma}{\sqrt{n}}$ is a 99% confidence interval for the unknown parameter θ.

However, using normal distribution is assuming the estimated variance sigma is the actual variance, which is not the case. If we use the estimated variance, the data actually follows a t-distribution which is slightly different from a normal distribution.Also since, our n is less than 35, we have to use a t-distribution. Instead of $\frac{\sqrt{n}(\bar{x}-\theta)}{\sigma}$ lying between -2.576 and 2.576, in a t-distribution with 12 degrees of freedom (degrees of freedom is calculated by n-1, in this case it is n is 13 therefore it is 12) $\frac{\sqrt{n}(\bar{x}-\theta)}{\sigma}$ will lie between -2.681 and 2.681 as indicated on the t-distribution table.

Calculation

Using the background information, we can conduct the exploration. All the calculations were done using Microsoft Excel. The highlighted cells are the predicted values

Table 1.Exponential Function for number of road accidents on National Highways

Year	ti	Yi	zi = ln(Yi)	ti^2	zi*ti	zi*ti^2		^Y = 29.2741e^-0.0011	(Yi-^Y)^2	(Yi-^Y)^2/^Y		Predicted Y	ti	Predicted ^
2005	-6	29.6	3.387774	36	-20.3266	121.9599		29.46971618	0.016974	0.000576		2018	7	29.04752
2006	-5	30.4	3.414443	25	-17.0722	85.36107		29.43702295	0.927325	0.031502		2019	8	29.0153
2007	-4	29	3.367296	16	-13.4692	53.87673		29.40436598	0.163512	0.005561		2020	9	28.98311
2008	-3	28.5	3.349904	9	-10.0497	30.14914		29.37174524	0.75994	0.025873		2021	10	28.95095
2009	-2	29.3	3.377588	4	-6.75518	13.51035		29.33916069	0.001534	5.23E-05				
2010	-1	30	3.401197	1	-3.4012	3.401197		29.30661229	0.480787	0.016405				
2011	0	30.1	3.404525	0	0	0		29.2741	0.682111	0.023301				
2012	1	29.1	3.370738	1	3.370738	3.370738		29.24162378	0.020057	0.000686				
2013	2	28.1	3.33577	4	6.671539	13.34308		29.20918358	1.230288	0.04212				
2014	3	28.2	3.339322	9	10.01797	30.0539		29.17677938	0.954098	0.032701				
2015	4	28.4	3.346389	16	13.38556	53.54223		29.14441112	0.554148	0.019014				
2016	5	29.6	3.387774	25	16.93887	84.69436		29.11207877	0.238067	0.008178				
2017	6	30.4	3.414443	36	20.48666	122.9199		29.07978229	1.742975	0.059938				
	0		**43.89716**	**182**	**-0.2028**	**616.1826**				**0.265906**	Fit is good is because it less than tabulated valu			
sumziti/sur	-0.00111													
zn = A	3.376705													
a	29.27415													
mean of yi	29.28462													
sd of yi	0.814295													
l_n	28.67913													
u_n	29.89011													

$\widehat{Y_1}=29.2741e^{-0.00111\,t}$, 99% confidence interval: 28.67913< θ (population mean) <29.89011

χ2(Chi -square value): 0.265906

Table2. Exponential Function for number of Persons Killed on National Highways

Year	ti	Yi	zi = ln(Yi)	ti^2	zi*ti	zi*ti^2		^Y = 29.27	(Yi-^Y)^2	(Yi-^Y)^2/^Y		Predicted Y	ti	Predicted ^
2005	-6	37.3	3.618993	36	-21.714	130.2838		36.85101	0.201589	0.00547		2018	7	34.24552
2006	-5	37.7	3.62966	25	-18.1483	90.7415		36.64374	1.115689	0.030447		2019	8	34.0529
2007	-4	35.5	3.569533	16	-14.2781	57.11252		36.43763	0.879148	0.024127		2020	9	33.86137
2008	-3	35.6	3.572346	9	-10.717	32.15111		36.23268	0.400283	0.011048		2021	10	33.67091
2009	-2	36	3.583519	4	-7.16704	14.33408		36.02888	0.000834	2.32E-05				
2010	-1	36.1	3.586293	1	-3.58629	3.586293		35.82623	0.074949	0.002092				
2011	0	37.1	3.613617	0	0	0		35.62472	2.176451	0.061094				
2012	1	35.3	3.563883	1	3.563883	3.563883		35.42434	0.015461	0.000436				
2013	2	33.2	3.50255	4	7.0051	14.0102		35.22509	4.100996	0.116423				
2014	3	34.1	3.529297	9	10.58789	31.76368		35.02696	0.859258	0.024531				
2015	4	35	3.555348	16	14.22139	56.88557		34.82995	0.028918	0.00083				
2016	5	34.5	3.540959	25	17.7048	88.52398		34.63404	0.017966	0.000519				
2017	6	36	3.583519	36	21.50111	129.0067		34.43923	2.435991	0.070733				
	0		**46.44952**	**182**	**-1.02658**	**651.9633**				**0.347774**	Fit is good is because it less than tabulated valu			
sumziti/sur	-0.00564									`				
zn = a	3.57304													
a'	35.62472													
mean of yi	35.64615													
sd of yi	1.283625													
l_n	34.69168													
u_n	36.60063													

$\widehat{Y_2} = 35.62472e^{-0.00564t}$, χ2: 0.347774, 34.69168< θ <36.60063

Table3 Exponential Function for number of persons injured on national highways

Year	ti	Yi	zi = ln(Yi)	ti^2	zi*ti	zi*ti^2		^Y = 29.27	(Yi-^Y)^2	(Yi-^Y)^2/^Y		Predicted Y	ti	Predicted ^
2005	-6	31.3	3.443618	36	-20.6617	123.9703		30.49839	0.642576	0.021069		2018	7	29.43776
2006	-5	30.8	3.427515	25	-17.1376	85.68787		30.41546	0.147867	0.004862		2019	8	29.35771
2007	-4	30.2	3.407842	16	-13.6314	54.52547		30.33276	0.017626	0.000581		2020	9	29.27789
2008	-3	28.6	3.353407	9	-10.0602	30.18066		30.25029	2.723447	0.09003		2021	10	29.19828
2009	-2	29.6	3.387774	4	-6.77555	13.5511		30.16803	0.322663	0.010696				
2010	-1	31.3	3.443618	1	-3.44362	3.443618		30.08601	1.473781	0.048986				
2011	0	30.5	3.417727	0	0	0		30.0042	0.245817	0.008193				
2012	1	30.1	3.404525	1	3.404525	3.404525		29.92262	0.031465	0.001052				
2013	2	28.9	3.363842	4	6.727683	13.45537		29.84126	0.885963	0.029689				
2014	3	29.9	3.397858	9	10.19358	30.58073		29.76012	0.019567	0.000658				
2015	4	29.1	3.370738	16	13.48295	53.93181		29.6792	0.335469	0.011303				
2016	5	29.6	3.387774	25	16.93887	84.69436		29.5985	2.26E-06	7.63E-08				
2017	6	30.3	3.411148	36	20.46689	122.8013		29.51802	0.611497	0.020716				
	0		**44.21739**	**182**	**-0.49554**	**620.2271**				**0.247834**	Fit is good is because it less than tabulated valu			
	Standard deviation													
b = sumziti/	-0.00272									`				
zn = a	3.401337													
a'	30.0042													
mean of yi	30.01538													
sd of yi	0.852297													
l_n	29.38164													
u_n	30.64913													

$\widehat{Y_3} = 30.0042e^{-0.00272t}$, χ2 (Chi Square Value) = 0.247834, 29.38164<θ<30.64913

Table4 Exponential function for number of road accidents on State Highways

Year	ti	Yi	zi = ln(Yi)	ti^2	zi*ti	zi*ti^2		^Y = 29.27	(Yi-^Y)^2	(Yi-^Y)^2/^Y		Predicted Y	ti	Predictrd ^
2005	-6	23.6	3.161247	36	-18.9675	113.8049		22.61716	0.965967	0.042709		2018	7	25.96223
2006	-5	18.5	2.917771	25	-14.5889	72.94427		22.85842	18.99579	0.83102		2019	8	26.23917
2007	-4	24.4	3.194583	16	-12.7783	51.11333		23.10224	1.684176	0.072901		2020	9	26.51905
2008	-3	25.6	3.242592	9	-9.72778	29.18333		23.34867	5.068493	0.217078		2021	10	26.80193
2009	-2	23.8	3.169686	4	-6.33937	12.67874		23.59772	0.040916	0.001734				
2010	-1	24.5	3.198673	1	-3.19867	3.198673		23.84944	0.423234	0.017746				
2011	0	24.6	3.202746	0	0	0		24.10383	0.246182	0.010213				
2012	1	24.2	3.186353	1	3.186353	3.186353		24.36094	0.025903	0.001063				
2013	2	25.6	3.242592	4	6.485185	12.97037		24.6208	0.958841	0.038944				
2014	3	25.2	3.226844	9	9.680532	29.0416		24.88342	0.100223	0.004028				
2015	4	24	3.178054	16	12.71222	50.84886		25.14885	1.319849	0.052481				
2016	5	25.3	3.230804	25	16.15402	80.77011		25.4171	0.013713	0.00054				
2017	6	25	3.218876	36	19.31325	115.8795		25.68822	0.47365	0.018438				
	0		**41.37082**	**182**	**1.931074**	**575.62**				**1.308897**	Fit is good is because it less than tabulated valu			
sumziti/sur	0.01061									`				
zn = A	3.182371													
a	24.10383													
mean of yi	24.17692													
sd of yi	1.828093													
l_n	22.8176													
u_n	25.53625													

$\widehat{Y_4} = 24.10383e^{0.01061t}$, χ2 = 1.308897, 22.8176<θ<25.53625

ISSN: 2456–5474 RNI No.UPBIL/2016/68367 Vol.-6* Issue-2* March- 2021
Innovation The Research Concept

Table5 Exponential function for number of persons killed on state highways

Year	ti	Yi	zi = ln(Yi)	ti^2	zi*ti	zi*ti^2		^Y = 29.27	(Yi-^Y)^2	(Yi-^Y)^2/^Y		Predicted Y	ti	Predicted ^
2005	-6	27.2	3.303217	36	-19.8193	118.9158		27.34655	0.021478	0.000785		2018	7	28.19387
2006	-5	26.8	3.288402	25	-16.442	82.21005		27.41082	0.373098	0.013611		2019	8	28.26012
2007	-4	27.7	3.321432	16	-13.2857	53.14292		27.47523	0.050521	0.001839		2020	9	28.32654
2008	-3	28.4	3.346389	9	-10.0392	30.1175		27.5398	0.739946	0.026868		2021	10	28.3931
2009	-2	27.1	3.299534	4	-6.59907	13.19813		27.60452	0.254538	0.009221				
2010	-1	27.3	3.306887	1	-3.30689	3.306887		27.66939	0.136447	0.004931				
2011	0	27.4	3.310543	0	0	0		27.73441	0.11183	0.004032				
2012	1	27.3	3.306887	1	3.306887	3.306887		27.79959	0.249586	0.008978				
2013	2	29.6	3.387774	4	6.775549	13.5511		27.86491	3.010521	0.10804				
2014	3	29.1	3.370738	9	10.11221	30.33664		27.9304	1.367971	0.048978				
2015	4	28	3.332205	16	13.32882	53.31527		27.99603	1.57E-05	5.62E-07				
2016	5	27.9	3.328627	25	16.64313	83.21567		28.06182	0.026187	0.000933				
2017	6	26.9	3.292126	36	19.75276	118.5165		28.12777	1.507416	0.053592				
	0		**43.19476**	**182**	**0.427197**	**603.1334**				**0.281809**	Fit is good is because it less than tabulated valu			
sumziti/sur	0.002347									`				
zn = A	3.322674													
a	27.73441													
mean of yi	27.74615													
sd of yi	0.84815													
l_n	27.11549													
u_n	28.37682													

$\widehat{Y_5} = 27.73441e^{0.002347t}$, $\chi 2 = 0.281809$, $27.11549<\theta<28.37682$

Table6 Exponential function for number of persons injured on state highways

Year	ti	Yi	zi = ln(Yi)	ti^2	zi*ti	zi*ti^2		^Y = 29.27	(Yi-^Y)^2	(Yi-^Y)^2/^Y		Predicted Y	ti	Predicted ^
2005	-6	25.7	3.246491	36	-19.4789	116.8737		25.94897	0.061986	0.002389		2018	7	26.32118
2006	-5	24.9	3.214868	25	-16.0743	80.3717		25.97741	1.160822	0.044686		2019	8	26.35003
2007	-4	26.2	3.265759	16	-13.063	52.25215		26.00589	0.037679	0.001449		2020	9	26.37892
2008	-3	27.5	3.314186	9	-9.94256	29.82767		26.0344	2.147996	0.082506		2021	10	26.40783
2009	-2	25.5	3.238678	4	-6.47736	12.95471		26.06293	0.316894	0.012159				
2010	-1	26	3.258097	1	-3.2581	3.258097		26.0915	0.008373	0.000321				
2011	0	26.1	3.261935	0	0	0		26.1201	0.000404	1.55E-05				
2012	1	25.9	3.254243	1	3.254243	3.254243		26.14873	0.061868	0.002366				
2013	2	27.6	3.317816	4	6.635632	13.27126		26.1774	2.023801	0.077311				
2014	3	26.8	3.288402	9	9.865206	29.59562		26.20609	0.352728	0.01346				
2015	4	26.3	3.269569	16	13.07828	52.3131		26.23482	0.004249	0.000162				
2016	5	25.8	3.250374	25	16.25187	81.25936		26.26357	0.2149	0.008182				
2017	6	25.4	3.234749	36	19.4085	116.451		26.29236	0.79631	0.030287				
	0		**42.41517**	**182**	**0.199389**	**591.6826**				**0.275293**	Fit is good is because it less than tabulated valu			
sumziti/sur	0.001096									`				
zn = A	3.262705													
a	26.1201													
mean of yi	26.13077													
sd of yi	0.781435													
l_n	25.54971													
u_n	26.71183													

$\widehat{Y_6} = 26.1201e^{0.001096t}$, $\chi 2 = 0.275293$, $25.54971<\theta<26.71183$

Table7 Modified Exponential Function for number of road accidents on National Highways

Year	ti	Yi	zi = ln(Yi)	ti^2	zi*ti	zi*ti^2		^Y = a*b^t	(Yi-^Y)^2	(Yi-^Y)^2/^Y		Predicted Y	ti	Predicted ^
2005	-6	29.6	3.387774	36	-20.3266	121.9599		29.47052	0.016764	0.000569		2018	7	29.0467
2006	-5	30.4	3.414443	25	-17.0722	85.36107		29.4377	0.926016	0.031457		2019	8	29.01435
2007	-4	29	3.367296	16	-13.4692	53.87673		29.40492	0.163959	0.005576		2020	9	28.98204
2008	-3	28.5	3.349904	9	-10.0497	30.14914		29.37217	0.760684	0.025898		2021	10	28.94976
2009	-2	29.3	3.377588	4	-6.75518	13.51035		29.33946	0.001557	5.31E-05				
2010	-1	30	3.401197	1	-3.4012	3.401197		29.30679	0.480545	0.016397				
2011	0	30.1	3.404525	0	0	0		29.27415	0.68203	0.023298				
2012	1	29.1	3.370738	1	3.370738	3.370738		29.24155	0.020036	0.000685				
2013	2	28.1	3.33577	4	6.671539	13.34308		29.20898	1.229841	0.042105				
2014	3	28.2	3.339322	9	10.01797	30.0539		29.17645	0.95346	0.032679				
2015	4	28.4	3.346389	16	13.38556	53.54223		29.14396	0.553477	0.018991				
2016	5	29.6	3.387774	25	16.93887	84.69436		29.1115	0.238629	0.008197				
2017	6	30.4	3.414443	36	20.48666	122.9199		29.07908	1.744822	0.060003				
	0		**43.89716**	**182**	**-0.2028**	**616.1826**				**0.265908**	Fit is good is because it less than tabulated valu			
sumziti/sur	-0.00111	0.998886	<- e^b											
zn = A	3.376705													
a	29.27415													
mean of yi	29.28462													
sd of yi	0.814295													
l_n	28.67913													
u_n	29.89011													

$\widehat{Y_A} = 29.27415(0.998886)^t$, $\chi 2 = 0.265908$, $28.67913<\theta<29.89011$

Table8 Modified Exponential Function for number of persons killed on national highways

Year	ti	Yi	zi = ln(Yi)	ti^2	zi*ti	zi*ti^2		^Y = a*b^t	(Yi-^Y)^2	(Yi-^Y)^2/^Y		Predicted Y	ti	Predicted ^
2005	-6	37.3	3.618993	36	-21.714	130.2838		36.90975	0.152292	0.004126		2018	7	34.01721
2006	-5	37.7	3.62966	25	-18.1483	90.7415		36.67877	1.042904	0.028433		2019	8	33.80433
2007	-4	35.5	3.569533	16	-14.2781	57.11252		36.44924	0.901053	0.024721		2020	9	33.59278
2008	-3	35.6	3.572346	9	-10.717	32.15111		36.22114	0.385814	0.010652		2021	10	33.38256
2009	-2	36	3.583519	4	-7.16704	14.33408		35.99447	3.06E-05	8.5E-07				
2010	-1	36.1	3.586293	1	-3.58629	3.586293		35.76922	0.109419	0.003059				
2011	0	37.1	3.613617	0	0	0		35.54537	2.416868	0.067994				
2012	1	35.3	3.563883	1	3.563883	3.563883		35.32293	0.000526	1.49E-05				
2013	2	33.2	3.50255	4	7.0051	14.0102		35.10188	3.617145	0.103047				
2014	3	34.1	3.529297	9	10.58789	31.76368		34.88221	0.611856	0.017541				
2015	4	34	3.526361	16	14.10544	56.42177		34.66392	0.44079	0.012716				
2016	5	34.5	3.540959	25	17.7048	88.52398		34.44699	0.00281	8.16E-05				
2017	6	36	3.583519	36	21.50111	129.0067		34.23143	3.127857	0.091374				
	0		**46.42053**	**182**	**-1.14253**	**651.4995**				**0.36376**	Fit is good is because it less than tabulated valu			
sumziti/sur	-0.00628	0.993742												
zn = A	3.57081													
a	35.54537													
mean of yi	35.56923													
sd of yi	1.353628													
l_n	34.56271													
u_n	36.57576													

$\widehat{Y_B} = 35.5437(0.993742)^t$, χ2 = 0.36376, 34.56271<θ<36.57576

Table9 Modified Exponential Function for number of persons injured on national highways

Year	ti	Yi	zi = ln(Yi)	ti^2	zi*ti	zi*ti^2		^Y = a*b^t	(Yi-^Y)^2	(Yi-^Y)^2/^Y		Predicted Y	ti	Predicted ^
2005	-6	31.3	3.443618	36	-20.6617	123.9703		30.49839	0.642576	0.021069		2018	7	29.43776
2006	-5	30.8	3.427515	25	-17.1376	85.68787		30.41546	0.147867	0.004862		2019	8	29.35771
2007	-4	30.2	3.407842	16	-13.6314	54.52547		30.33276	0.017626	0.000581		2020	9	29.27789
2008	-3	28.6	3.353407	9	-10.0602	30.18066		30.25029	2.723447	0.09003		2021	10	29.19828
2009	-2	29.6	3.387774	4	-6.77555	13.5511		30.16803	0.322663	0.010696				
2010	-1	31.3	3.443618	1	-3.44362	3.443618		30.08601	1.473781	0.048986				
2011	0	30.5	3.417727	0	0	0		30.0042	0.245817	0.008193				
2012	1	30.1	3.404525	1	3.404525	3.404525		29.92262	0.031465	0.001052				
2013	2	28.9	3.363842	4	6.727683	13.45537		29.84126	0.885963	0.029689				
2014	3	29.9	3.397858	9	10.19358	30.58073		29.76012	0.019567	0.000658				
2015	4	29.1	3.370738	16	13.48295	53.93181		29.6792	0.335469	0.011303				
2016	5	29.6	3.387774	25	16.93887	84.69436		29.5985	2.26E-06	7.63E-08				
2017	6	30.3	3.411148	36	20.46689	122.8013		29.51802	0.611497	0.020716				
	0		**44.21739**	**182**	**-0.49554**	**620.2271**	Predicted			**0.247834**	Fit is good is because it less than tabulated valu			
sumziti/sur	-0.00272	0.997281												
zn = A	3.401337													
a	30.0042													
mean of yi	30.01538													
sd of yi	0.852297													
l_n	29.38164													
u_n	30.64913													

$\widehat{Y_C} = 30.0042(0.997281)^t$, χ2= 0.247834, 29.38164<θ<30.64913

Table10 Modified Exponential function for number of road accidents on State Highways

Year	ti	Yi	zi = ln(Yi)	ti^2	zi*ti	zi*ti^2		^Y = a*b^t	(Yi-^Y)^2	(Yi-^Y)^2/^Y		Predicted Y	ti	Predicted ^
2005	-6	23.6	3.161247	36	-18.9675	113.8049		22.61716	0.965967	0.042709		2018	7	25.96223
2006	-5	18.5	2.917771	25	-14.5889	72.94427		22.85842	18.99579	0.83102		2019	8	26.23917
2007	-4	24.4	3.194583	16	-12.7783	51.11333		23.10224	1.684176	0.072901		2020	9	26.51905
2008	-3	25.6	3.242592	9	-9.72778	29.18333		23.34867	5.068493	0.217078		2021	10	26.80193
2009	-2	23.8	3.169686	4	-6.33937	12.67874		23.59772	0.040916	0.001734				
2010	-1	24.5	3.198673	1	-3.19867	3.198673		23.84944	0.423234	0.017746				
2011	0	24.6	3.202746	0	0	0		24.10383	0.246182	0.010213				
2012	1	24.2	3.186353	1	3.186353	3.186353		24.36094	0.025903	0.001063				
2013	2	25.6	3.242592	4	6.485185	12.97037		24.6208	0.958841	0.038944				
2014	3	25.2	3.226844	9	9.680532	29.0416		24.88342	0.100223	0.004028				
2015	4	24	3.178054	16	12.71222	50.84886		25.14885	1.319849	0.052481				
2016	5	25.3	3.230804	25	16.15402	80.77011		25.4171	0.013713	0.00054				
2017	6	25	3.218876	36	19.31325	115.8795		25.68822	0.47365	0.018438				
	0		**41.37082**	**182**	**1.931074**	**575.62**	Predicted			**1.308897**	Fit is good is because it less than tabulated valu			
sumziti/sur	0.01061	1.010667												
zn = A	3.182371													
a	24.10383													
mean of yi	24.17692													
sd of yi	1.828093													
l_n	22.8176													
u_n	25.53625													

$\widehat{Y_D} = 24.10383(1.010667)^t$, χ2=1.308897, 22.8176< θ<25.53625

Table11 Modified Exponential function for number of persons killed on state highways

Year	ti	Yi	zi = ln(Yi)	ti^2	zi*ti	zi*ti^2		^Y = a*b^t	(Yi-^Y)^2	(Yi-^Y)^2/^Y		Predicted Y	ti	
2005	-6	27.2	3.303217	36	-19.8193	118.9158		27.34655	0.021478	0.000785		2018	7	28.19387
2006	-5	26.8	3.288402	25	-16.442	82.21005		27.41082	0.373098	0.013611		2019	8	28.26012
2007	-4	27.7	3.321432	16	-13.2857	53.14292		27.47523	0.050521	0.001839		2020	9	28.32654
2008	-3	28.4	3.346389	9	-10.0392	30.1175		27.5398	0.739946	0.026868		2021	10	28.3931
2009	-2	27.1	3.299534	4	-6.59907	13.19813		27.60452	0.254538	0.009221				
2010	-1	27.3	3.306887	1	-3.30689	3.306887		27.66939	0.136447	0.004931				
2011	0	27.4	3.310543	0	0	0		27.73441	0.11183	0.004032				
2012	1	27.3	3.306887	1	3.306887	3.306887		27.79959	0.249586	0.008978				
2013	2	29.6	3.387774	4	6.775549	13.5511		27.86491	3.010521	0.10804				
2014	3	29.1	3.370738	9	10.11221	30.33664		27.9304	1.367971	0.048978				
2015	4	28	3.332205	16	13.32882	53.31527		27.99603	1.57E-05	5.62E-07				
2016	5	27.9	3.328627	25	16.64313	83.21567		28.06182	0.026187	0.000933				
2017	6	26.9	3.292126	36	19.75276	118.5165		28.12777	1.507416	0.053592				
	0		**43.19476**	**182**	**0.427197**	**603.1334**	Predicted			**0.281809**	Fit is good is because it less than tabulated valu			
sumziti/sur	0.002347	1.00235												
zn = A	3.322674													
a	27.73441													
mean of yi	27.74615													
sd of yi	0.84815													
l_n	27.11549													
u_n	28.37682													

$\widehat{Y_E} = 27.73441(1.00235)^t$, χ2=0.281809, 27.11549<θ<28.37682

Table12 Modified Exponential function for number of persons injured on state highways

Year	ti	Yi	zi = ln(Yi)	ti^2	zi*ti	zi*ti^2		^Y = a*b^t	(Yi-^Y)^2	(Yi-^Y)^2/^Y		Predicted Y	ti	Predicted ^
2005	-6	25.7	3.246491	36	-19.4789	116.8737		25.94897	0.061986	0.002389		2018	7	26.32118
2006	-5	24.9	3.214868	25	-16.0743	80.3717		25.97741	1.160822	0.044686		2019	8	26.35003
2007	-4	26.2	3.265759	16	-13.063	52.25215		26.00589	0.037679	0.001449		2020	9	26.37892
2008	-3	27.5	3.314186	9	-9.94256	29.82767		26.0344	2.147996	0.082506		2021	10	26.40783
2009	-2	25.5	3.238678	4	-6.47736	12.95471		26.06293	0.316894	0.012159				
2010	-1	26	3.258097	1	-3.2581	3.258097		26.0915	0.008373	0.000321				
2011	0	26.1	3.261935	0	0	0		26.1201	0.000404	1.55E-05				
2012	1	25.9	3.254243	1	3.254243	3.254243		26.14873	0.061868	0.002366				
2013	2	27.6	3.317816	4	6.635632	13.27126		26.1774	2.023801	0.077311				
2014	3	26.8	3.288402	9	9.865206	29.59562		26.20609	0.352728	0.01346				
2015	4	26.3	3.269569	16	13.07828	52.3131		26.23482	0.004249	0.000162				
2016	5	25.8	3.250374	25	16.25187	81.25936		26.26357	0.2149	0.008182				
2017	6	25.4	3.234749	36	19.4085	116.451		26.29236	0.79631	0.030287				
	0		**42.41517**	**182**	**0.199389**	**591.6826**				**0.275293**	Fit is good is because it less than tabulated valu			
sumziti/sur	0.001096	1.001096												
zn = A	3.262705													
a	26.1201													
mean of yi	26.13077													
sd of yi	0.781435													
l_n	25.54971													
u_n	26.71183													

$\widehat{Y_F} = 26.1201(1.001096)^t$, χ2=0.275293, 25.549713<θ<26.71182

Findings

Table 13 Exponential Function

Years	National Highways			State Highways		
	Road Accidents	Persons Killed	Persons Injured	Persons Killed	Persons Killed	Persons Injured
2018	29.04752	34.24552	29.43776	25.96223	28.19387	26.32118
2019	29.0153	34.0529	29.35771	26.23917	28.26012	26.35003
2020	28.98311	33.86137	29.27789	26.51905	28.32654	26.37892
2021	28.95095	33.67091	29.19828	26.80193	28.3931	26.40783

Table14 Modified Exponential Function

Years	National Highways			State Highways		
	Road Accidents	Road Accidents	Persons Injured	Persons Killed	Persons Injured	Persons Killed
2018	29.0467	34.01721	29.43776	25.96223	28.19387	26.32118
2019	29.01435	33.80433	29.35771	26.23917	28.26012	26.35003
2020	28.98204	33.59278	29.27789	26.51905	28.32654	26.37892
2021	28.94976	33.38256	29.19828	26.80193	28.3931	26.40783

Graphs

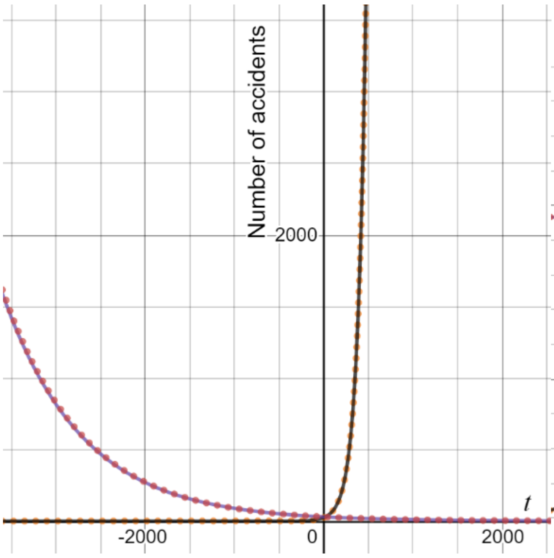

Figure 1: Graphs from very far.

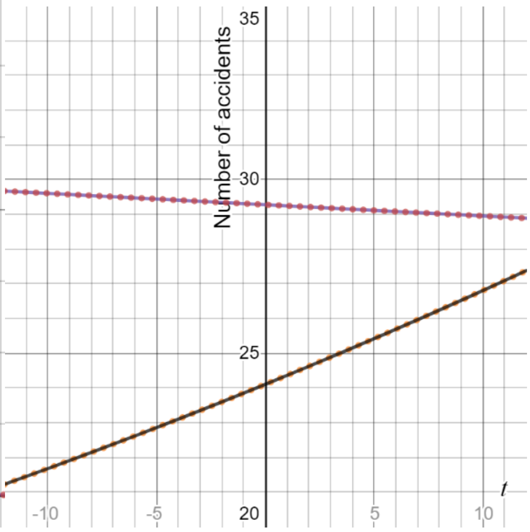

Figure 2: Graphs from close up.

Figure 1 is shown to show how the lines are exponential and modified exponential curves. Figure 2 is the graph at a realistic t value.

The blue line is the exponential function on number of road accidents national highways, and the dotted line is the respective modified exponential function for the number of road accidents on national highways. The black line is the exponential function for road accidents on state highways and the orange dotted line is the respective modified exponential function. At a glance, they look like the same line, however, they do have minute differences as calculated above. The national highways curves both show a decaying trend, meaning that the predicted values should decrease in value, which they do. However, the state highway curves show an increasing trend. This is very concerning as the curves are both for number of road accidents. This could mean that the government must focus more on prevention of road accidents in state highways.

The distributions also seem very linear, indicating that a linear regression might have worked as well, although it may have not worked as well as the exponential and modified exponential functions.

Aim of the Study

It was part of his study curriculum. The aim was delope a suitable pre dictor so that we could predict about traffic situation on high ways in the country.

Conclusion

We can conclude that for the purpose of prediction of number of accidents, number of deaths, and number of injuries, persons, the above given exponential model works quite satisfactorily as tested by Chi square test of goodness of fit. For the fit to be considered good, the calculated chi squared value must be less than the tabulated value (18.3 for a 5% margin of error) for corresponding degrees of freedom and percentage level of significance. In this case, the calculated valuesare much lower than the tabulated value, indicating a very good fit.

Similarly, we can conclude that the modified exponential model also works quite satisfactorily as tested by Chi square test of goodness of fit.

Although both functions are well below the tabulated chi square value and their chi square values are almost same, the exponential model works marginally better as its chi-square values are lower by small margin. The difference is so small that in reality both models can be used for prediction.

Annex

Year	National Highways			State Highways			Other Roads		
	Road Accidents	Persons Killed	Persons Injured	Road Accidents	Persons Killed	Persons Injured	Road Accidents	Persons Killed	Persons Injured
2005	29.6	37.3	31.3	23.6	27.2	25.7	46.8	35.5	43.0
2006	30.4	37.7	30.8	18.5	26.8	24.9	51.1	35.5	44.3
2007	29.0	35.5	30.2	24.4	27.7	26.2	46.6	36.8	43.6
2008	28.5	35.6	28.6	25.6	28.4	27.5	45.9	36.0	43.9
2009	29.3	36.0	29.6	23.8	27.1	25.5	46.9	36.9	44.9
2010	30.0	36.1	31.3	24.5	27.3	26.0	45.5	36.6	42.7
2011	30.1	37.1	30.5	24.6	27.4	26.1	45.3	35.5	43.4
2012	29.1	35.3	30.1	24.2	27.3	25.9	46.7	37.4	44.0
2013	28.1	33.2	28.9	25.6	29.6	27.6	46.3	37.2	43.5
2014	28.2	34.1	29.9	25.2	29.1	26.8	46.6	36.8	43.3
2015	28.4	35.0	29.1	24.0	28.0	26.3	47.6	37.0	44.6
2016	29.6	34.5	29.6	25.3	27.9	25.8	45.1	37.6	44.6

Figure 3: Data from 2005 to 2017

References

1. *"Degrees of Freedom in Statistics". Statistics by Jim, 12, Nov. 2018, statisticsbyjim.com/hypothesis-esting/degrees-freedom-statistics/. Accessed 15, Feb. 2020.*
1. *Goon, Atindra M., et al. Fundamentals of Statistics. 1968.*
2. *Gupta, S. C., and V. K. Kapoor. Fundamentals of Mathematical Statistics. 1980.*
3. *India Environment Portal | News, Reports, Documents, Blogs, Data, Analysis on Environment &Development India, South Asia,*
4. *www.indiaenvironmentportal.org.in/files/file/road%20accidents%20in%20India%202017.pdf. Accessed 4 July 2019.*
5. *"Interval Estimation". UBC Wiki, wiki.ubc.ca/ Interval Estimation. Accessed 15 Feb. 2020.*
6. *San Jose State University, www.sjsu.edu/faculty/gerstman/StatPrimer/t-table.pdf.*
7. *Siegel, Sidney. Nonparametric Statistics for the Behavioral Sciences. 1959.*