

VIDYASAGAR UNIVERSITY



ANALYSIS OF DATA ON DEATH BY ROAD ACCIDENT IN INDIA : A COMPARATIVE STUDY

Registration No: - 1160429 of 2020-2021

Roll No: - 1126116-200132

Session: - 2022-2023

Acknowledgment:

The success and final outcome of this project required a lot of guidance and assistance from many people and I am extremely privileged to have got this all along with the completion of my project. All that I have done is only due to their supervision and assistance and I would not forget to thank them. I respect and thank, for providing me an opportunity to do the project work under the department of Statistics, Haldia Government College and giving u all the support and guidance that I required, which made me complete the project duly on time. I owe my deep gratitude to our project guides Dr. Shyamsundar Sahoo, Mr. Sibsankar Karan, Mr. Tanmay Kumar Maity and Mr. Bijitesh Halder, who took keen interest on my project work and guided me all along, till the completion of our project work by providing all the necessary information for developing a good system. I would not forget to remember my parents, and also my friends for their encouragement and more over for their timely support and guidance till the completion of our project work.

Contents:

❖ Introduction:.....	4
❖ Objective:.....	4
❖ Data:.....	4
❖ Methodology:.....	5-6
❖ Analysis & result.....	7-33
❖ Conclusion:.....	34
❖ Reference:.....	34

1. Introduction:

Road accidents occur on daily basis and the severity of injuries caused by it is enormous. According to the World Health Organizing (WHO), road traffic injuries are the sixth leading cause of death in India. Road accidents scenario in the country is a matter of great concern, as it became a major social, economical and health problem. India has nearly six crore motor vehicles on the road. Over 85000 people are killed annually in our country. People aged between 15 and 44 years account for 48 percent of global road traffic deaths. About 1.25 million people die each year as result of traffic crashes. Road traffic injuries are the leading cause of death among young people aged 15-29 years. 90 percent of the worlds fatalities on the roads occur in low-and middle income countries, even though these countries have approximately half of the worlds vehicle. Half of those dying on the world's roads are vulnerable road users: pedestrian, cyclist and motor cyclists. Kerala is the one of the leading states in high rate of accidents and injuries. Main reason for this is the increasing number of vehicles, is not paving way for broader roads and road safety measures. The main causes of road accidents are:

-  Fault of driver
-  Fault of cyclist
-  Fault of passenger except
-  Bad weather
-  Bad road
-  Technical defect
-  Fault of pedestrian
-  Drunken driving
-  Others

2. Objective:

1. In this project we compare the rate of person killed in each accident by different types of vehicles is same or different
2. We want to compare death by accident between south and north India through different ways (vehicle, wheather,etc.).
3. We want to analysis which type of cause for accident is more responsible for more death.
4. We compare to the rate of person killed in each accident in various season is same or different.

3. Data collection:

We collect the data from “Ministry of road Transport and Highways”.From this website we only take the data of road accident in India in 2021.

For the analysis we take the data of 50 citsy all over the India. There are only number of accident and number of death for a particular city.

4. Methodology:

To analysis the data we use some statistical methodology which are given below

1. Test for goodness of fit:

The chi-square test (Snedecor and Cochran, 1989) is used to test if a sample of data came from a population with a specific distribution.

An attractive feature of the chi-square goodness-of-fit test is that it can be applied to any univariate distribution for which you can calculate the cumulative distribution function. The chi-square goodness-of-fit test is applied to binned data (i.e., data put into classes). This is actually not a restriction since for non-binned data you can simply calculate a histogram or frequency table before generating the chi-square test. However, the value of the chi-square test statistic are dependent on how the data is binned. Another disadvantage of the chi-square test is that it requires a sufficient sample size in order for the chi-square approximation to be valid. The chi-square test is an alternative to the Anderson-Darling and Kolmogorov-Smirnov goodness-of-fit tests. The chi-square goodness-of-fit test can be applied to discrete distributions such as the binomial and the Poisson.

The chi-square test is defined for the hypothesis:

H₀: The data follow a specified distribution.

H_a: The data do not follow the specified distribution.

Test Statistic: For the chi-square goodness-of-fit computation, the data are divided into k bins and the test statistic is defined as

$$\chi^2 = \sum (O_i - E_i)^2 / E_i$$

where O_i is the observed frequency for bin i and E_i is the expected frequency for bin i. The expected frequency is calculated by

$$E_i = N(F(Y_u) - F(Y_l))$$

where F is the cumulative distribution function for the distribution being tested, Y_u is the upper limit for class i, Y_l is the lower limit for class i, and N is the sample size.

The test statistic follows, approximately, a chi-square distribution with (k - c) degrees of freedom where k is the number of non-empty cells and c = the number of estimated parameters (including location and scale parameters and shape parameters) for the distribution + 1. For example, for a 3-parameter Weibull distribution, c = 4.

Therefore, the hypothesis that the data are from a population with the specified distribution is rejected if

$$\chi^2 > \chi^2_{1-\alpha, k-c}$$

where $\chi^2_{1-\alpha, k-c}$ is the chi-square critical value with k - c degrees of freedom and significance level α .

2. Poisson Distribution:

A discrete random variable X said to follow poisson distribution if its pmf is in the form of

$$P(X=x) = \frac{e^{-\lambda} \lambda^x}{x!}; X=0,1,2,3,\dots$$

Where λ is a parameter which is greter than 0,

In notation , $X \sim \text{poission}(\lambda)$

3. Exact test for Poisson parameter of two population:

We may be interested in a comparison of the parameters λ_1 and λ_2 of 2 Poisson distribution. Let $x_{1j} (j = 1, 2, \dots, n_1)$ be a random sample from the first distribution and $x_{2i} (i = 1, 2, \dots, n_2)$ be a random sample from the second, the 2 samples themselves being mutually independent. A test for

Hypothesis Testing:

$$\begin{array}{lcl} & H_{10}: \lambda_1 = \lambda_2 & (\lambda_1 / \lambda_2 = 1) \\ \text{again} & H_{11}: \lambda_1 > \lambda_2 & \text{or } H_{11}: \lambda_1 < \lambda_2 \end{array}$$

let be denote, $y_1 = \sum_{i=1}^{n_1} x_{1i}$
 $y_2 = \sum_{i=1}^{n_2} x_{2i}$

And $y = y_1 + y_2$

Under null hypothesis, the distribution of above variable

$$f(y_1) = \exp(-n_1\lambda) (n_1\lambda)^{y_1} / y_1!$$

$$f(y_2) = \exp(-n_2\lambda) (n_2\lambda)^{y_2} / y_2!$$

$$f(y) = \exp[-(n_1 + n_2)\lambda] [(n_1 + n_2)\lambda]^y / y!$$

and conditional distribution of y_1 given y is

$$f(y_1|y) = \frac{\exp[-(n_1 + n_2)\lambda] (n_1\lambda)^{y_1} (n_2\lambda)^{y-y_1} / [y_1! (y-y_1)!]}{\exp[-(n_1 + n_2)\lambda] [(n_1 + n_2)\lambda]^y / y!}$$

which follows Binomial($y, n_1 / (n_1 + n_2)$)

Denoting the observed value of y_1 and y by y_{10} and y_0

a) in the case of $H_{11}: \lambda_1 > \lambda_2$

$$P[y_1 \geq y_{10} / y = y_0] = \sum_{y_1 \geq y_{10}} \binom{y_0}{y_1} \left(\frac{n_1}{n_1 + n_2} \right)^{y_1} \left(\frac{n_2}{n_1 + n_2} \right)^{y_0 - y_1}$$

And compare with α for acceptance or rejection of null hypothesis.

If $P[y_1 \geq y_{10} / y = y_0] > \alpha$, we accept null hypothesis.

b) in the case of $H_{11}: \lambda_1 < \lambda_2$

$$P[y_1 \leq y_{10} / y = y_0] = \sum_{y_1 \leq y_{10}} \binom{y_0}{y_1} \left(\frac{n_1}{n_1 + n_2} \right)^{y_1} \left(\frac{n_2}{n_1 + n_2} \right)^{y_0 - y_1}$$

And compare with α for acceptance or rejection of null hypothesis.

If $P[y_1 \leq y_{10} / y = y_0] > \alpha$, we accept null hypothesis.

5. Analysis & Result:

state	Four/six wheelar			two/three wheeler			other(pedestrian,etc)		
	numbe r of road accide nt	Kille d	no of killed in per accident(A)	numbe r of road accide nt	kille d	no of killed in per accident(B)	numbe r of road accide nt	kille d	no of killed in per accident(C)
Agra	202	132	0.653465	518	296	0.571429	195	121	0.620513
Ahmedabad	104	27	0.259615	917	216	0.235551	412	161	0.390777
Allahabad (Prayagraj)	447	211	0.472036	431	188	0.436195	287	136	0.473868
Amritsar	10	5	0.5	62	43	0.693548	26	17	0.653846
Asansol Durgapur	71	39	0.549296	136	107	0.786765	191	156	0.816754
Aurangabad	74	9	0.121622	278	104	0.374101	119	41	0.344538
Bengaluru	498	70	0.140562	1826	340	0.186199	889	244	0.274466
Bhopal	255	22	0.086275	1852	171	0.092333	509	60	0.117878
Chandigarh	38	7	0.184211	125	64	0.512	45	25	0.555556
Coimbatore	68	6	0.088235	579	160	0.276339	219	68	0.310502
Chennai	432	67	0.155093	3242	628	0.193708	1360	303	0.222794
Delhi	276	72	0.26087	2234	552	0.24709	2210	615	0.278281
Dhanbad	34	31	0.911765	96	57	0.59375	57	35	0.614035
Faridabad	36	17	0.472222	280	96	0.342857	189	98	0.518519
Ghaziabad	176	75	0.426136	549	272	0.495446	99	48	0.484848
Gwalior	613	139	0.226754	473	73	0.154334	744	147	0.197581
Hyderabad	283	22	0.077739	1329	177	0.133183	661	98	0.14826
Indore	692	61	0.08815	1846	257	0.13922	1138	166	0.14587
Jabalpur	1115	190	0.170404	667	124	0.185907	2073	153	0.073806
Jaipur	898	175	0.194878	790	267	0.337975	477	183	0.383648
Jamshedpur	32	11	0.34375	69	24	0.347826	51	30	0.588235
Jodhpur	169	50	0.295858	312	135	0.432692	112	52	0.464286
Kannur	650	12	0.018462	486	70	0.144033	38	40	1.052632
Kanpur	346	195	0.563584	517	206	0.398453	491	197	0.401222
Khozikode	726	15	0.020661	651	95	0.145929	0	21	0
Kochi	852	70	0.08216	884	71	0.080317	45	0	0
Kolkata	359	19	0.052925	732	95	0.129781	626	82	0.13099
Kollam	334	16	0.047904	889	133	0.149606	329	35	0.106383
Kota	37	5	0.135135	243	45	0.185185	88	25	0.284091
Lucknow	380	129	0.339474	325	99	0.304615	85	33	0.388235
Ludhiana	100	75	0.75	280	218	0.778571	98	87	0.887755
Madurai	62	8	0.129032	375	99	0.264	181	47	0.259669
Mallapuram	1168	29	0.024829	956	185	0.193515	23	77	3.347826
Meerut	291	115	0.395189	316	118	0.373418	204	128	0.627451
Mumbai	684	171	0.25	462	42	0.090909	1084	174	0.160517
Nagpur	116	24	0.206897	641	174	0.271451	201	70	0.348259

Nashik	43	8	0.186047	303	124	0.409241	124	53	0.427419
Patna	113	38	0.336283	119	70	0.588235	152	102	0.671053
Pune	93	24	0.258065	449	148	0.329621	199	83	0.417085
Raipur	391	38	0.097187	1030	328	0.318447	342	106	0.309942
Rajkot	33	17	0.515152	208	70	0.336538	105	49	0.466667
Srinagar	66	3	0.045455	35	6	0.171429	230	31	0.134783
Surat	60	23	0.383333	415	136	0.327711	229	113	0.49345
Thiruvanthapuram	322	28	0.086957	782	46	0.058824	334	43	0.128743
Thrissur	124	8	0.064516	1256	137	0.109076	339	56	0.165192
Tiruchirapalli	35	5	0.142857	242	77	0.318182	122	48	0.393443
Vadodra	74	10	0.135135	285	96	0.336842	105	40	0.380952
Varanasi	146	99	0.678082	135	68	0.503704	142	69	0.485915
Vijaywada city	120	17	0.141667	710	157	0.221127	398	113	0.28392
Vizaq	376	21	0.055851	1449	242	0.167012	514	105	0.20428

Four/six wheeler: truck, bus ,car.

Two/three wheeler: motore cycle,auto riksha.

Other: pedestrian,other non motorised vehicle,bicycle.

1.1. **Fitting of Poisson distribution(number of person killed in 5 accident):**

X_i be a random variable which denote the number of person killed in 5 accident. Assume that X follow Poisson(λ) distribution. The parameter λ is to be estimated through method of moment. So $\hat{\lambda} = (\sum_{i=1}^n X_i)/n$

As the random variable is only takes values $X=0,1,3,4,\dots$

So I manipulate the data through this method given below,

If $0 \leq X \leq 0.5$ then X take value 0.

If $0.5 < X \leq 1.5$ then X take value 1.

If $1.5 < X \leq 2.5$ then X take value 2.

And so on.....

So new data set,

X	Four /six wheeler(A)	two/three wheelar(B)	Others(C)
0	15	4	3
1	19	21	17
2	9	17	18
3	4	6	8
4	2	2	2
>5	0	0	2

We fit Poisson distribution on this data set .And test goodness of fit through this fitting model.

1.1.a: **Fitting Poisson distribution in population A(Four/six wheeler):**

Here,

$$\hat{\lambda}_A = (\sum_{i=1}^n X_i)/50 = 1.16$$

Observed frequency(x) = $P(X=x)*50$

$$= \frac{e^{-1.16} 0.16^x}{x!} * 50$$

Table:

	A(lamda=1.16)		
X	observed(Oi)	expected(Ei)	chi square component
0	15	16	0.065
1	19	18	0.0556
2	9	11	0.3636
3	4	4	0
4	2	1	1
total	50	50	1.4817

Goodness of fit:

The chi-square test is defined for the hypothesis:

H₀: The data follow a poisson distribution.

H_a: The data do not follow the poisson distribution.

Test Statistic: For the chi-square goodness-of-fit computation, the data are divided into k bins and the test statistic is defined as

$$\chi^2 = \sum_{i=1}^k (O_i - E_i)^2 / E_i$$

here $k=5$, Under H₀ $\chi^2 \sim \chi^2_{(k-1)}$

we conduct the test under $\alpha=0.05$ level of significance

So, test statistics value $\chi^2 = 1.48$

Tabulated value, $\chi^2_{\alpha(k-1)} = 9.89$

p-value = 0.89

as $\chi^2 < \chi^2_{\alpha(k-1)}$, so we accept H₀. (i.e. population A comes from poisson distribution)

1.1.b :Fitting Poisson distribution in population B(two/three wheeler):

Here,

$$\hat{\lambda}_B = (\sum_{i=1}^n X_i) / 50 = 1.62$$

Observed frequency(x) = P(X=x)*50

$$= \frac{e^{-1.62} 1.62^x}{x!} * 50$$

Table:

	B(lamda=1.62)		
X	observed(Oi)	expected(Ei)	chi square component
0	4	10	3.6
1	21	16	1.5
2	17	13	1.23
3	6	7	0.14
4	2	3	0.333

5	0	1	1
total	50	50	7.8695

Goodness of fit:

here $k=6$, Under $H_0 \chi^2 \sim \chi^2_{(k-1)}$

we conduct the test under $\alpha=0.05$ level of significance

So, test statistics value, $\chi^2=7.895$

Tabulated value, $\chi^2_{\alpha(k-1)}=11.07$

p-value=0.16

as $\chi^2 < \chi^2_{\alpha(k-1)}$, so we accept H_0 . (i.e. population B comes from Poisson distribution)

1.1.c: Fitting Poisson distribution in population C(others):

Here,

$$\hat{\lambda}_C = (\sum_{i=1}^n X_i) / 50 = 1.7$$

Observed frequency(x) = $P(X=x) * 50$

$$= \frac{e^{-1.7} 1.7^x}{x!} * 50$$

Table:

	C(lamda=1.7)		
X	observed(Oi)	expected(Ei)	chi square component
0	3	9	4
1	17	16	0.0625
2	18	14	1.1429
3	8	7	1.1429
4	2	3	0.333
5	2	1	1
total	50	50	6.6816

Goodness of fit:

here $k=6$, Under $H_0 \chi^2 \sim \chi^2_{(k-1)}$

we conduct the test under $\alpha=0.05$ level of significance

So, test statistics value $\chi^2=6.6816$

Tabulated value, $\chi^2_{\alpha(k-1)}=11.07$

p-value=0.24

as $\chi^2 < \chi^2_{\alpha(k-1)}$, so we accept H_0 . (i.e. population C comes from Poisson distribution)

1.2. Test for Equality of Poisson parameter of three different type of vehicle:

Let us denote

X_{1i} : random variable that represent the number of person killed in 5 accident by four/six wheeler (A) in i th class.

X_{2i} : random variable that represent the number of person killed in 5 accident by two/three wheeler in i th class.

X_{3i} : random variable that represent the number of person killed in 5

accident by other(pedestrian) in i th class.

We can assume that $X_{ji}; j=1(1)3, i=1(1)50$, comes from Poisson distribution.

$$X_{1i} \sim \text{Poisson}(\lambda_1)$$

$$X_{2i} \sim \text{Poisson}(\lambda_2)$$

$$X_{3i} \sim \text{Poisson}(\lambda_3)$$

They are independent

1.2.a: Hypothesis Testing:

i) $H_{10}: \lambda_1 = \lambda_2 = \lambda_3$

against H_{11} : at least one of Poisson parameter is not equal.

For testing of this hypothesis we use large sample test.

As $n=50$

Large sample test:

Assume $Y_1 = \sum_{i=1}^{50} X_{1i}$
 $Y_2 = \sum_{i=1}^{50} X_{2i}$
 $Y_3 = \sum_{i=1}^{50} X_{3i}$

So, $Y_1 \sim \text{poission}(50\lambda_1)$
 $Y_2 \sim \text{poission}(50\lambda_2)$
 $Y_3 \sim \text{poission}(50\lambda_3)$

From large sample theory,

Under H_0 ,

$$\chi^2 = \sum_{i=1}^3 \left(\frac{Y_i - n\hat{\lambda}_i}{\sqrt{n\hat{\lambda}_i}} \right)^2 \text{ follow } \chi^2_{(k-1)}.$$

Here $n=50, k=3$;

$$\hat{\lambda}_i = \hat{\lambda} = \frac{\sum_{k=1}^3 \sum_{i=1}^{50} X_{ki}}{150}; \text{ for all } i$$

$$= 1.4933$$

we conduct the test under $\alpha=0.05$ level of significance

So, test statistics value $\chi^2 = 2.966$

Tabulated value, $\chi^2_{0.05, 2} = 5.99$

as $\chi^2 < \chi^2_{\alpha(k-1)}$, so we accept H_0 . (i.e. population A,B,C comes from same Poisson parameter)

$$\text{so, } \lambda_1 = \lambda_2 = \lambda_3$$

1.3. Comparison between north and south india in death rate by accident by different vehicle:

	North India								
	Four/six wheeler			Two/Three wheeler			other(pedestrian,etc)		
city	number of road accident	killed	No of killed in total 100 accident	number of road accident	killed	no of killed in total 100 accident	number of road accident	killed	no of killed in total 100 accident
Agra	202	132	0.65	518	296	0.57	195	121	0.62
Allahabad(Prayagraj)	447	211	0.47	431	188	0.43	287	136	0.47
Amritsar	10	5	0.5	62	43	0.69	26	17	0.65
Chandigarh	38	7	0.18	125	64	0.51	45	25	0.55
Delhi	276	72	0.26	2234	552	0.24	2210	615	0.27
Faridabad	36	17	0.47	280	96	0.34	189	98	0.51
Ghaziabad	176	75	0.42	549	272	0.49	99	48	0.48
Jaipur	898	175	0.19	790	267	0.33	477	183	0.38
Jodhpur	169	50	0.29	312	135	0.43	112	52	0.46
Kannur	650	12	0.01	486	70	0.14	38	40	1.05
Kanpur	346	195	0.56	517	206	0.39	491	197	0.4
Kota	37	5	0.13	243	45	0.18	88	25	0.28
Lucknow	380	129	0.33	325	99	0.3	85	33	0.38
Ludhiana	100	75	0.75	280	218	0.77	98	87	0.88
Meerut	291	115	0.39	316	118	0.37	204	128	0.62
Srinagar	66	3	0.04	35	6	0.17	230	31	0.13
Varanasi	146	99	0.67	135	68	0.5	142	69	0.48

south india									
	four wheeler			two/three wheeler			other(pedestrian,etc)		
city	number of road accident	killed	no of killed in total 100 accident	number of road accident	killed	no of killed in total 100 accident	number of road accident	killed	no of killed in total 100 accident
Vizag	376	21	0.05	1449	242	0.16	514	105	0.2
Vijaywada city	120	17	0.14	710	157	0.22	398	113	0.28
Tiruchirapalli	35	5	0.14	242	77	0.31	122	48	0.39
Thrissur	124	8	0.06	1256	137	0.1	339	56	0.16
Thiruvanthapuram	322	28	0.08	782	46	0.05	334	43	0.12
Mallapuram	1168	29	0.02	956	185	0.19	23	77	3.34
Madurai	62	8	0.12	375	99	0.26	181	47	0.25
Kollam	334	16	0.04	889	133	0.14	329	35	0.1
Kochi	852	70	0.08	884	71	0.08	45	0	0
Khozikode	726	15	0.02	651	95	0.14	0	21	0
Hyderabad	283	22	0.07	1329	177	0.13	661	98	0.14
Chennai	432	67	0.15	3242	628	0.19	1360	303	0.22
Coimbatore	68	6	0.08	579	160	0.27	219	68	0.31
Bengaluru	498	70	0.14	1826	340	0.18	889	244	0.27
Mumbai	684	171	0.25	462	42	0.09	1084	174	0.16
Nagpur	116	24	0.2	641	174	0.27	201	70	0.34
Nashik	43	8	0.18	303	124	0.4	124	53	0.42
Pune	93	24	0.25	449	148	0.32	199	83	0.41

We want to compare the rate of death by accident between south and north India by different category of vehicles.

X_i ; $i=1(1)3$, be the random variable denote number of person died in 100 accident by i th type of vehicle in north India.

Y_i ; $i=1(1)3$, be the random variable denote number of person died in 100 accident by i th type of vehicle in south India.

We can assume that X_i comes from Poisson distribution.

$$X_{1i} \sim \text{Poisson}(\lambda_1)$$

$$X_{2i} \sim \text{Poisson}(\lambda_2)$$

$$X_{3i} \sim \text{Poisson}(\lambda_3)$$

They are independent

And also,

We can assume that Y_i comes from Poisson distribution.

$$Y_{1i} \sim \text{Poisson}(\lambda_a)$$

$$Y_{2i} \sim \text{Poisson}(\lambda_b)$$

$$Y_{3i} \sim \text{Poisson}(\lambda_c)$$

They are independent.

	North India		
X	X1	X2	X3
0	2	0	0
1	5	5	3
2	6	8	7

3	3	3	5
4	1	1	1
5	0	0	1

	South India		
Y	Y1	Y2	Y3
0	9	4	3
1	9	11	9
2	0	3	5
3	0	0	0
4	0	0	0
>5	0	0	1

Hypothesis Testing:

- i) $H_{10}: \lambda_1 = \lambda_a$
 against $H_{11}: \lambda_1 > \lambda_a$
 ii) $H_{10}: \lambda_2 = \lambda_b$
 against $H_{11}: \lambda_2 > \lambda_b$
 iii) $H_{10}: \lambda_3 = \lambda_c$
 against $H_{11}: \lambda_3 > \lambda_c$

1.3.a. Exact Test for two Poisson parameter:

We are interested to compare poisson parameter λ_i

We know $\hat{\lambda}_i = \bar{X}$

- i) $H_{10}: \lambda_1 = \lambda_a$ **against** $H_{11}: \lambda_1 > \lambda_a$

Here $n_1=17, n_2=18$

```
> x<-poisson.test(c(30,9),c(17,18),alternative = "greater")
> x
```

Comparison of Poisson rates
 data: c(30, 9) time base: c(17, 18)
 count1 = 30, expected count1 = 18.943, p-value = 0.000282
 alternative hypothesis: true rate ratio is greater than 1
 95 percent confidence interval:
 1.817219 Inf
 sample estimates:
 rate ratio : 3.529412

we conduct the test under 0.05% level of significance

we accept alternative hypothesis. So $\lambda_1 > \lambda_a$

i.e. the rate of person killed by accident through for wheeler in north india is greater than accident cause by four wheeler in south india.

- ii) $H_{20}: \lambda_2 = \lambda_b$ **against** $H_{21}: \lambda_2 > \lambda_b$

```
> poisson.test(c(34,17),c(17,18),alternative = "greater")
```

Comparison of Poisson rates

data: c(34, 17) time base: c(17, 18)
count1 = 34, expected count1 = 24.771, p-value = 0.006963
alternative hypothesis: true rate ratio is greater than 1
95 percent confidence interval:
1.257725 Inf
sample estimates:
rate ratio
2.117647

we conduct the test under 0.05% level of significance

we reject Null hypothesis. So $\lambda_2 > \lambda_b$

i.e. the rate of person killed by accident in north india through two/three wheeler is greater than accident in south india cause by two/three wheeler.

iii) $H_{30}: \lambda_3 = \lambda_c$ against $H_{31}: \lambda_3 > \lambda_c$

```
> poisson.test(c(41,24),c(17,18),alternative = "greater")
```

Comparison of Poisson rates

data: c(41, 24) time base: c(17, 18)
count1 = 41, expected count1 = 31.571, p-value = 0.01313
alternative hypothesis: true rate ratio is greater than 1
95 percent confidence interval:
1.154322 Inf
sample estimates:
rate ratio
1.808824

we conduct the test under 0.05% level of significance

we reject Null hypothesis. So $\lambda_3 > \lambda_c$

i.e. the rate of person killed by accident in north india cause by pedestrian and other is greater than compare to accident in south India.

2. Accident due to different cause:

over speeding			drunken driving/driving in wrong side			jumping red line/use mobile phone		
total accident	killed	total killed in 5 accident	total accident	killed	total killed in 5 accident	total accident	killed	total killed in 5 accident
419	251	2.995	45	35	3.889	32	21	3.28125
1,338	401	1.499	93	3	0.161	2	0	0
257	124	2.412	406	181	2.229	365	173	2.369863
65	40	3.077	27	19	3.519	6	6	5
179	141	3.939	3	1	1.667	1	1	5
427	129	1.511	9	1	0.556	0	0	0
3,041	601	0.988	101	27	1.337	17	3	0.882353
2,211	216	0.488	0	0	0.000	0	0	0
184	83	2.255	11	5	2.273	13	8	3.076923
729	179	1.228	27	4	0.741	1	0	0
4,880	952	0.975	143	44	1.538	0	0	0
1,824	437	1.198	374	81	1.083	489	132	1.349693
104	61	2.933	24	13	2.708	0	0	0
505	211	2.089	0	0	0.000	0	0	0
347	155	2.233	80	49	3.063	17	13	3.823529
1,316	261	0.992	306	51	0.833	0	0	0
1,976	249	0.630	47	3	0.319	5	2	2
3,674	484	0.659	2	0	0.000	0	0	0
1,380	301	1.091	2259	123	0.272	0	0	0
1,911	565	1.478	90	20	1.111	10	4	2
108	51	2.361	44	14	1.591	0	0	0
390	159	2.038	20	7	1.750	0	0	0
649	76	0.586	125	5	0.200	8	1	0.625
222	87	1.959	181	94	2.597	221	98	2.217195
786	76	0.483	41	4	0.488	8	0	0
898	61	0.340	28	3	0.536	8	0	0
10	12	6.000	62	9	0.726	6	3	2.5
786	78	0.496	24	4	0.833	4	0	0
357	73	1.022	8	1	0.625	0	0	0
298	109	1.829	197	66	1.675	162	51	1.574074
345	289	4.188	26	35	6.731	107	56	2.616822
559	146	1.306	48	8	0.833	11	0	0
1,873	239	0.638	82	11	0.671	1	0	0
264	92	1.742	229	118	2.576	79	36	2.278481
2,191	381	0.869	39	6	0.769	0	0	0
123	46	1.870	56	9	0.804	4	0	0

over speeding			drunken driving/driving in wrong side			jumping red line/use mobile phone		
448	177	1.975	20	8	2.000	2	0	0
191	105	2.749	45	22	2.444	20	11	2.75
0	6	0.000	18	6	1.667	0	0	0
1,430	439	1.535	55	0	0.000	20	0	0
335	129	1.925	4	0	0.000	0	0	0
331	40	0.604	0	0	0.000	0	0	0
692	272	1.965	12	0	0.000	0	0	0
848	56	0.330	46	2	0.217	4	0	0
1,300	135	0.519	44	4	0.455	6	0	0
353	108	1.530	15	5	1.667	3	2	3.333333
426	134	1.573	21	4	0.952	5	1	1
76	40	2.632	111	71	3.198	76	45	2.960526
1,101	267	1.213	127	20	0.787	0	0	0
1,785	261	0.731	83	7	0.422	60	4	0.333333

From the given data we want to check whether or not the rate of person killed in each accident through three different population (i.e. over speeding, drunken driving, using mobile phone) same or not .

2.1.Fitting of poisson distribution(number of person killed in 5 accident):

X_i be a random variable which denote the number of person killed in $5i$ accident. Assume that X follow poisson(λ) distribution.The parameter λ is to be estimated through method of moment.So $\hat{\lambda}=(\sum_{i=1}^n Xi)/n$

As the random variable is only takes values $X=0,1,3,4,\dots$

So I manipulate the data through this method given below,

If $0 \leq X \leq 0.5$ then X take value 0.

If $0.5 < X \leq 1.5$ then X take value 1.

If $1.5 < X \leq 2.5$ then X take value 2.

And so on.....

So the new data set

x	X_1	X_2	X_3
0	6	15	30
1	19	16	4
2	17	11	7
3	5	5	6
4	2	2	1
>5	1	1	2

I fit Poisson distribution on this data set .And test goodness of fit through this fitting model.

2.1.a: Fitting Poisson distribution in population X_1 (Over speeding):

Here, $\hat{\lambda}_A=(\sum_{i=1}^n Xi)/50$
 $=1.62$

$$\text{Observed frequency}(x) = P(X=x) * 50$$

$$= \frac{e^{-1.62} 1.62^x}{x!} * 50$$

Table:

	X ₁ (lamda=1.62)		
	observed(Oi)	expected(Ei)	chi square component
X			
0	6	10	1.6
1	19	16	0.5625
2	17	13	1.23
3	5	7	0.5714
4	2	3	0.333
>5	1	1	0
total	50	50	4.298

Goodness of fit:

The chi-square test is defined for the hypothesis:

H₀: The data follow a poisson distribution.

H_a: The data do not follow the poisson distribution.

Test Statistic: For the chi-square goodness-of-fit computation, the data are divided into k bins and the test statistic is defined as

$$\chi^2 = \sum_{i=1}^k (O_i - E_i)^2 / E_i$$

here $k=6$, Under H₀ $\chi^2 \sim \chi^2_{(k-1)}$

we conduct the test under $\alpha=0.05$ level of significance

So, test statistics value $\chi^2 = 0.4298$

Tabulated value, $\chi^2_{\alpha(k-1)} = 11.070$

p-value = 0.507

as $\chi^2 < \chi^2_{\alpha(k-1)}$, so we accept H₀. (i.e. population X₁ comes from poisson distribution)

2.1.b :Fitting poisson distribution in population B(drunken driving/driving wrong side):

Here,

$$\hat{\lambda}_B = (\sum_{i=1}^n Xi) / 50$$

$$= 1.32$$

$$\text{Observed frequency}(x) = P(X=x) * 50$$

$$= \frac{e^{-1.32} 1.32^x}{x!} * 50$$

Table:

	X ₂ (lamda=1.32)		
X	observed(Oi)	expected(Ei)	chi square component
0	15	13	0.3

1	16	17	0.058
2	11	12	0.833
3	5	5	0
4	2	2	0
>5	1	1	0
total	50	50	0.4498

Goodness of fit:

here $k=6$, Under $H_0 \quad \chi^2 \sim \chi^2_{(k-1)}$

we conduct the test under $\alpha=0.05$ level of significance

So, test statistics value $\chi^2=12.8$

Tabulated value, $\chi^2_{\alpha(k-1)}=11.070$

p-value=0.9938

as $\chi^2 < \chi^2_{\alpha(k-1)}$, so we accept H_0 . (i.e. population X_2 comes from poisson distribution)

2.1.c:Fitting poisson distribution in population X_3 (using mobile phone/jumping red line):

Here,

$$\hat{\lambda}_c = (\sum_{i=1}^n Xi)/50$$

=1

Observed frequency(x) = $P(X=x)*50$

$$= \frac{e^{-1}1^x}{x!} * 50$$

Table:

X	C(lamda=1)		
	observed(Oi)	expected(Ei)	chi square component
0	30	18	8
1	4	18	10.889
2	7	9	0.444
3	6	3	3
4	1	1	0
>5	2	1	1
total	50	50	23.33

Goodness of fit:

here $k=6$, Under $H_0 \quad \chi^2 \sim \chi^2_{(k-1)}$

we conduct the test under $\alpha=0.05$ level of significance

So, test statistics value $\chi^2=23.333$

Tabulated value, $\chi^2_{\alpha(k-1)}=11.070$

p-value=0.00029

as $\chi^2 > \chi^2_{\alpha(k-1)}$, so we reject H_0 . (i.e. population X_3 does not comes from poisson distribution).

Comparison of Poisson rates

data: c(81, 66) time base: c(50, 50)

count1 = 81, expected count1 = 73.5, p-value = 0.1241

alternative hypothesis: true rate ratio is greater than 1

95 percent confidence interval:

0.9224447 Inf

sample estimates:

rate ratio

1.227273

As p- value > 0.05 so we accept null.

2.3. Comparison between north and south india in death rate by accident due to different causes:

Data:

	no of person killed in each accident(North India)	
city	over speeding	drunken driving/driving wrong side
Agra	0.585516179	3.889
Allahabad(Prayagraj)	0.604229607	2.229
Amritsar	1.022408964	3.519
Chandigarh	1.197916667	2.273
Delhi	1.478283621	1.083
Faridabad	1.742424242	0.000
Ghaziabad	1.82885906	3.063
Jaipur	1.959459459	1.111
Kannur	2.038461538	0.200
Kanpur	2.089108911	2.597
Kota	2.233429395	0.625
Lucknow	2.255434783	1.675
Ludhiana	2.412451362	6.731
Varanasi	2.631578947	3.198
Meerut	2.99522673	2.576
Srinagar	3.076923077	0.000
Jodhpur	4.188405797	1.750

	no of person killed in each accident(south india)	
city	over speeding	drunken driving/driving wrong side
Vizag	0	0.422
Tiruchirapalli	0.330188679	1.667
Pune	0.339643653	1.667
Nashik	0.48346056	2.000

Mumbai	0.496183206	0.769
Nagpur	0.519230769	0.804
Kochi	0.630060729	0.536
Vijaywada city	0.638013881	0.787
Coimbatore	0.731092437	0.741
Chennai	0.869465997	1.538
Hyderabad	0.975409836	0.319
Thiruvanthapuram	0.988161789	0.217
Thrissur	1.21253406	0.455
Madurai	1.227709191	0.833
Mallapuram	1.305903399	0.671
Kollam	1.529745042	0.833
Khozikode	1.869918699	0.488
Bengaluru	1.975446429	1.337

We want to compare the rate of death by accident due to different causes between south and north india .

X_i ; $i=1(1)2$, be the random variable denote number of person died in a accident due to ith type of cause in north india.

Y_i ; $i=1(1)2$, be the random variable denote number of person died in a accident due to ith type of cause in south india.

We can assume that X_i comes from poisson distribution.

$$X_{1i} \sim \text{poisson}(\lambda_1)$$

$$X_{2i} \sim \text{poisson}(\lambda_2)$$

They are independent

And also,

We can assume that Y_i comes from poisson distribution.

$$Y_{1i} \sim \text{poisson}(\lambda_a)$$

$$Y_{2i} \sim \text{poisson}(\lambda_b)$$

They are independent.

	north india	
x	X1	X2
0	1	3
1	4	3
2	8	4
3	3	4
4	1	2
>5	0	1

	south India	
Y	Y1	Y2
0	5	5
1	10	9
2	3	4
3	0	0
4	0	0
>5	0	0

Hypothesis Testing:

- | | |
|---------|---------------------------------|
| i) | $H_{10}: \lambda_1 = \lambda_a$ |
| against | $H_{11}: \lambda_1 > \lambda_a$ |
| ii) | $H_{10}: \lambda_2 = \lambda_b$ |
| against | $H_{11}: \lambda_2 > \lambda_b$ |

2.3.a.Fisher Exact Test for two poisson parameter:

We are interested to compare poisson parameter λ_i

We know $\hat{\lambda}_i = \bar{X}$

i) $H_{10}: \lambda_1 = \lambda_a$ **against** $H_{11}: \lambda_1 > \lambda_a$

`poisson.test(c(33,16),c(17,18),alternative = "greater")`

Comparison of Poisson rates

data: c(33, 16) time base: c(17, 18)
count1 = 33, expected count1 = 23.8, p-value = 0.006172
alternative hypothesis: true rate ratio is greater than 1
95 percent confidence interval:
1.279548 Inf
sample estimates:
rate ratio
2.183824

we conduct the test under 0.05% level of significance

we accept Null hypothesis. So $\lambda_1 = \lambda_a$

i.e. the rate of person killed by accident due to overspeeding in north india equal to accident in south india.

ii) $H_{20}: \lambda_2 = \lambda_b$ **against** $H_{21}: \lambda_2 > \lambda_b$

`poisson.test(c(36,17),c(17,18),alternative = "greater")`

Comparison of Poisson rates

data: c(36, 17) time base: c(17, 18)
count1 = 36, expected count1 = 25.743, p-value = 0.003464
alternative hypothesis: true rate ratio is greater than 1
95 percent confidence interval:
1.339747 Inf
sample estimates:
rate ratio
2.242215

we conduct the test under 0.05% level of significance

we reject Null hypothesis. So $\lambda_2 > \lambda_b$

i.e. the rate of person killed by accident due to drunken driving/driving wrong side in north india is greater than to accident in south india.

3. Accident in different weather:

	sunny	rainy	foggy	hail/sleet	other
City	no of killed in each accident	no of killed in each accident	no of killed in each accident	no of killed in each accident	no of killed in each accident
Agra	0.60321716	0.62195122	0.624277457	0.555555556	0.55625
Ahmedabad	0.28142458	1	0	0	0
Allahabad(Prayagraj)	0.48518519	0.446902655	0.436936937	0.450980392	0.469135802
Amritsar	0.45454545	0	0.75	0	0.771929825
Asansol Durgapur	0.66233766	0.75	0.729166667	0	0.845744681
Aurangabad	0.33185841	0.2	0.222222222	0	0
Bengaluru	0.2023849	0.222222222	0.220779221	0	0
Bhopal	0.09769335	0	0	0	0.091358025
Chandigarh	0.45049505	1	0.5	0	0
Coimbatore	0.26740947	0.333333333	0.268292683	0	0.5
Chennai	0.1983197	0.160857909	0.31372549	0	0.25
Delhi	0.23556942	0.271794872	0.25	0	0.304093567
Dhanbad	0.66666667	0.505494505	0.625	0.75	1
Faridabad	0.42450766	0	0.354166667	0	0
Ghaziabad	0.45987654	0.509803922	0.4375	0	0.553956835
Gwalior	0.19690064	0	0	0	0.195088677
Hyderabad	0.11377246	0.25	0.154639175	0.076923077	0.139534884
Indore	0.13499851	0	0	0	0.095846645
Jabalpur	0.13014504	0.104294479	0.474747475	0.308333333	0.022397892
Jaipur	0.2889096	0.263157895	0	0	0
Jamshedpur	0.49295775	0.348837209	0.166666667	0	0.6
Jodhpur	0.40140845	1	0.655172414	0	0.335766423
Kannur	0.09532888	0.180952381	0.15	0	0
Kanpur	0.43005181	0.456140351	0.451697128	0.504273504	0.418367347
Khozikode	0.08860759	0.1875	0.102564103	0	0.153061224
Kochi	0.07285181	0.154639175	0.115384615	0	0
Kolkata	0.11089239	0.262135922	0	0	0
Kollam	0.10458992	0.133333333	0.1875	0	0.21875
Kota	0.20380435	0	0	0	0
Lucknow	0.36875	0.348066298	0.317535545	0.286821705	0.321100917
Ludhiana	0.92519685	0	0.559139785	0.62745098	0.7625
Madurai	0.24959481	0	0	0	0
Mallapuram	0.13264129	0.145962733	0.153846154	0	0
Meerut	0.46317829	0.5	0.4125	0	0.411764706
Mumbai	0.1704918	0.1875	0	0	0
Nagpur	0.27974948	0	0	0	0
Nashik	0.38841202	1	0	0	0
Patna	0.49324324	0.574468085	0.589473684	0	0
Pune	0.34412955	0	0	0	0

Raipur	0.31014904	0.102409639	0.045454545	0	0
Rajkot	0.36760125	0.647058824	0.875	0	0
Srinagar	0.1358885	0.111111111	0	0	0
Surat	0.38095238	0.28125	0.454545455	0	0.401315789
Thiruvanthapuram	0.08208955	0	0.16	0	0.042857143
Thrissur	0.11229947	0.147982063	0	0	0
Tiruchirapalli	0.33773087	0.105263158	0	0	0
Vadodra	0.30296128	0.52	0	0	0
Varanasi	0.56462585	0.694736842	0.643678161	0.4	0.25
Vijaywada city	0.22461815	0.260869565	0.382352941	0	0.310344828
Vizaq	0.1553766	0.128787879	0.224719101	0.428571429	0

From the given data we want to check whether or not the rate of person killed in each accident through different population (sunny,rainy,foggy,hail/sleet,others) same or not .

3.1.Fitting of poisson distribution(in each 5 accident):

X_i be a random variable which denote the number of person killed in 5 accident.Assume that X follow poisson(λ) distribution.The parameter λ is to be estimated through method of moment.So $\hat{\lambda}=(\sum_{i=1}^n X_i)/n$

As the random variable is only takes values $X=0,1,3,4,\dots$

So I manipulate the data through this method given below,

If $0 \leq X \leq 0.5$ then X take value 0.

If $0.5 < X \leq 1.5$ then X take value 1.

If $1.5 < X \leq 2.5$ then X take value 2.

And so on.....

So new data set,

x	x1	x2	x3	x4	x5
0	5	11	17	41	28
1	21	21	13	1	6
2	19	6	11	4	9
3	4	7	6	3	3
4	0	1	3	1	3
>5	1	1	0	0	1
total	50	50	50	50	50

I fit poisson distribution on this data set .And test goodness of fit through this fitting model.

3.1.a: Fitting poisson distribution in population X_1 (sunny weather):

Here,

$$\hat{\lambda}_A = (\sum_{i=1}^n X_i)/50 \\ = 1.52$$

Observed frequency(x) = $P(X=x)*50$

$$= \frac{e^{-1.52} 1.52^x}{x!} * 50$$

Table:

	sunny(lamda=1.52)		
	observed(Oi)	expected(Ei)	chi square component
X1			
0	5	11	3.27
1	21	17	0.9
2	19	13	2.78
3	4	6	0.66
4	0	2	2
>5	1	1	0
total	50	50	9.6498

Goodness of fit:

The chi-square test is defined for the hypothesis:

H₀: The data follow a poisson distribution.

H_a: The data do not follow the poisson distribution.

Test For the chi-square goodness-of-fit computation, the data are divided into k bins

Statistic: and the test statistic is defined as

$$\chi^2 = \sum_{i=1}^k (O_i - E_i)^2 / E_i$$

here $k=6$, Under H₀ $\chi^2 \sim \chi^2_{(k-1)}$

we conduct the test under $\alpha=0.05$ level of significance

So, test statistics value $\chi^2 = 9.64$

Tabulated value, $\chi^2_{\alpha(k-1)} = 11.070$

p-value = 0.085

as $\chi^2 < \chi^2_{\alpha(k-1)}$, so we accept H₀. (i.e. population X₁ comes from poisson distribution)

3.1.b :Fitting Poisson distribution in population x2(rainy weather):

Here,

$$\hat{\lambda}_B = (\sum_{i=1}^n X_i) / 50$$

$$= 1.26$$

Observed frequency(x) = P(X=x)*50

$$= \frac{e^{-1.26} 1.26^x}{x!} * 50$$

Table:

	rainy(lamda=1.26)		
	observed(Oi)	expected(Ei)	chi square component
X2			
0	11	14	0.6
1	21	18	0.5
2	6	11	2.2
3	7	5	0.8
4	1	1	0

>5	1	1	0
total	50	50	4.2156

Goodness of fit:

here $k=6$, Under $H_0 \chi^2 \sim \chi^2_{(k-1)}$

we conduct the test under $\alpha=0.05$ level of significance

So, test statistics value $\chi^2=4.22$

Tabulated value, $\chi^2_{\alpha(k-1)}=11.07$

p-value=0.595

as $\chi^2 < \chi^2_{\alpha(k-1)}$, so we accept H_0 . (i.e. population X_2 comes from poisson distribution)

3.1.c: Fitting poisson distribution in population X_3 (foggy weather):

Here,

$$\hat{\lambda}_c = (\sum_{i=1}^n Xi) / 50 \\ = 0.67$$

Observed frequency(x) = $P(X=x) * 50$

$$= \frac{e^{-0.68} 0.68^x}{x!} * 50$$

Table:

foggy(lamda=0.68)			
X3	observed(Oi)	expected(Ei)	chi square component
0	17	14	0.6429
1	13	18	1.388
2	11	12	0.08
3	6	5	0.2
4	3	1	4
total	50	50	6.3151

Goodness of fit:

here $k=5$, Under $H_0 \chi^2 \sim \chi^2_{(k-1)}$

we conduct the test under $\alpha=0.05$ level of significance

So, test statistics value $\chi^2=6.3151$

Tabulated value, $\chi^2_{\alpha(k-1)}=9.488$

p-value=0.172

as $\chi^2 < \chi^2_{\alpha(k-1)}$, so we accept H_0 . (i.e. population X_3 comes from poisson distribution).

3.1.d: Fitting poisson distribution in population X_4 (hail weather):

Here,

$$\hat{\lambda}_d = (\sum_{i=1}^n Xi) / 50 \\ = 0.44$$

Observed frequency(x) = $P(X=x) * 50$

$$= \frac{e^{-0.44} 0.44^x}{x!} * 50$$

Table:

	hail(lamda=0.44)		
X4	observed(Oi)	expected(Ei)	chi square component
0	41	32	2.5
1	1	13	11.07
2	4	3	0.3
3	3	1	4
4	1	1	0
total	50	50	17.9415

Goodness of fit:

here k=5, Under $H_0 \quad \chi^2 \sim \chi^2_{(k-1)}$

we conduct the test under $\alpha=0.05$ level of significance

So, test statistics value $\chi^2 = 17.94$

Tabulated value, $\chi^2_{\alpha(k-1)} = 9.488$

p-value=0.001

as $\chi^2 > \chi^2_{\alpha(k-1)}$, so we reject H_0 . (i.e. population X_4 doesnot come from poisson distribution).

3.1.e:Fitting poisson distribution in population X_5 (other weather):

Here,

$$\hat{\lambda}_e = \left(\sum_{i=1}^n Xi \right) / 50$$

=1

Observed frequency(x) = $P(X=x) * 50$

$$= \frac{e^{-1} 1^x}{x!} * 50$$

Table:

	other(lamda=1)		
X5	observed(Oi)	expected(Ei)	chi square component
0	28	18	5.5
1	6	18	8
2	9	9	0
3	3	3	0
4	3	1	4
>5	1	1	0
total	50	50	17.5556

Goodness of fit:

here k=6, Under $H_0 \quad \chi^2 \sim \chi^2_{(k-1)}$

we conduct the test under $\alpha=0.05$ level of significance

So, test statistics value $\chi^2 = 17.22$

Tabulated value, $\chi^2_{\alpha(k-1)} = 11.070$

p-value = 0.003

as $\chi^2 > \chi^2_{\alpha(k-1)}$, so we reject H_0 . (i.e. population X_5 does not come from poisson distribution).

3.2. Test for Equality of poisson parameter of three different cause for accident:

Let us denote

X_{1i} : random variable that represent the number of person killed in a accident in clear weather (x_1) in i th class.

X_{2i} : random variable that represent the number of person killed in a accident in rainy weather (x_2) in i th class.

X_{3i} : random variable that represent the number of person killed in a accident in foggy weather (x_3) in i th class.

As the remaining population does not come from poisson distribution, so we do not use in testing problem.

We can assume that $X_{ji}; j=1(1)3, i=1(1)50$, comes from poisson distribution.

$$X_{1i} \sim \text{poisson}(\lambda_1)$$

$$X_{2i} \sim \text{poisson}(\lambda_2)$$

$$X_{3i} \sim \text{poisson}(\lambda_3)$$

They are independent

3.2.a: Hypothesis Testing:

i) $H_{10}: \lambda_1 = \lambda_2 = \lambda_3$

against H_{11} : atleast one of Poisson parameter is not equal.

For testing of this hypothesis we use large sample test.

As $n=50$

Large sample test:

Assume

$$Y_1 = \sum_{i=1}^{50} X_{1i}$$

$$Y_2 = \sum_{i=1}^{50} X_{2i}$$

$$Y_3 = \sum_{i=1}^{50} X_{3i}$$

So,

$$Y_1 \sim \text{poisson}(50\lambda_1)$$

$$Y_2 \sim \text{poisson}(50\lambda_2)$$

$$Y_3 \sim \text{poisson}(50\lambda_3)$$

From large sample theory,

$$\chi^2 = \sum_{i=1}^3 \left(\frac{Y_i - n\hat{\lambda}_i}{\sqrt{n\hat{\lambda}_i}} \right)^2 \text{ follow } \chi^2_{(k-1)}.$$

Here $n=50, k=3$;

Under H_0 $\hat{\lambda}_i = \hat{\lambda} = \frac{\sum_{k=1}^3 \sum_{i=1}^{50} X_{ki}}{150}$; for all i

$$=1.36$$

we conduct the test under $\alpha=0.05$ level of significance

So, test statistics value $\chi^2=1.44$

Tabulated value, $\chi^2_{0.05,2}=5.991$

as $\chi^2 < \chi^2_{\alpha(k-1)}$, so we accept H_0 . (i.e. population x_1, x_2, x_3 comes from same poisson parameter)
so, $\lambda_1 = \lambda_2 = \lambda_3$

3.3. Comparison between north and south india in death rate by accident due to different weather:

	north India		
city	no of killed in each accident(sunny)X1	no of killed in each accident(rainy)X2	no of killed in each accident(foggy)X3
Agra	0.60321716	0.62195122	0.624277457
Allahabad(Prayagraj)	0.48518519	0.44690265	0.436936937
Amritsar	0.45454545	0	0.75
Chandigarh	0.45049505	1	0.5
Delhi	0.23556942	0.27179487	0.25
Faridabad	0.42450766	0	0.354166667
Ghaziabad	0.45987654	0.50980392	0.4375
Jaipur	0.2889096	0.26315789	0
Kannur	0.09532888	0.18095238	0.15
Kanpur	0.43005181	0.45614035	0.451697128
Kota	0.20380435	0	0
Lucknow	0.36875	0.3480663	0.317535545
Ludhiana	0.92519685	0	0.559139785
Varanasi	0.56462585	0.69473684	0.643678161
Meerut	0.46317829	0.5	0.4125
Srinagar	0.1358885	0.11111111	0
Jodhpur	0.40140845	1	0.655172414

	south india		
city	no of killed in each accident(sunny)X1	no of killed in each accident(rainy)X2	no of killed in each accident(foggy)X3
Vizag	0.1553766	0.12878788	0.224719101
Tiruchirapalli	0.33773087	0.10526316	0
Pune	0.34412955	0	0
Nashik	0.38841202	1	0
Mumbai	0.1704918	0.1875	0

Nagpur	0.27974948	0	0
Kochi	0.07285181	0.15463918	0.115384615
Vijaywada city	0.22461815	0.26086957	0.382352941
Coimbatore	0.26740947	0.33333333	0.268292683
Chennai	0.1983197	0.16085791	0.31372549
Hyderabad	0.11377246	0.25	0.154639175
Thiruvanthapuram	0.08208955	0	0.16
Thrissur	0.11229947	0.14798206	0
Madurai	0.24959481	0	0
Mallapuram	0.13264129	0.14596273	0.153846154
Kollam	0.10458992	0.13333333	0.1875
Khozikode	0.08860759	0.1875	0.102564103
Bengaluru	0.2023849	0.22222222	0.220779221

We want to compare the rate of death by accident due to different weather between south and north India .

X_i ; $i=1(1)3$, be the random variable denote number of person died in 5 accident due to i th type of weather in north India.

Y_i ; $i=1(1)3$, be the random variable denote number of person died in 5 accident due to i th type of weather in south india.

We can assume that X_i comes from poisson distribution.

$$X_{1i} \sim \text{poisson}(\lambda_1)$$

$$X_{2i} \sim \text{poisson}(\lambda_2)$$

$$X_{3i} \sim \text{poisson}(\lambda_3)$$

They are independent

And also,

We can assume that Y_i comes from poisson distribution.

$$Y_{1i} \sim \text{poisson}(\lambda_a)$$

$$Y_{2i} \sim \text{poisson}(\lambda_b)$$

$$Y_{3i} \sim \text{poisson}(\lambda_c)$$

They are independent.

	north india		
x	x1	x2	x3
0	1	4	3
1	4	4	2
2	9	4	7
3	2	3	4
4	0	0	1
5	1	2	0

	south india		
y	y1	y2	y3
0	3	4	7
1	12	12	9
2	3	1	2
3	0	0	0
4	0	0	0
5	0	1	0

Hypothesis Testing:

- i) $H_{10}: \lambda_1 = \lambda_a$
 against $H_{11}: \lambda_1 > \lambda_a$
 ii) $H_{10}: \lambda_2 = \lambda_b$

against $H_{11}: \lambda_2 > \lambda_b$
 iii) $H_{10}: \lambda_3 = \lambda_c$
 against $H_{11}: \lambda_3 > \lambda_c$

3.3.a. Exact Test for two Poisson parameter:

We are interested to compare poisson parameter λ_i

We know $\hat{\lambda}_i = \bar{X}$

i) $H_{10}: \lambda_1 = \lambda_a$ **against** $H_{11}: \lambda_1 > \lambda_a$

Here $n_1=17, n_2=18$

> `poisson.test(c(33,18),c(17,18),alternative = "greater")`

Comparison of Poisson rates

data: c(33, 18) time base: c(17, 18)

count1 = 33, expected count1 = 24.771, p-value = 0.01487

alternative hypothesis: true rate ratio is greater than 1

95 percent confidence interval:

1.159947 Inf

sample estimates:

rate ratio

1.941176

we conduct the test under 0.05% level of significance

we accept alternative hypothesis. So $\lambda_1 > \lambda_a$

i.e. the rate of person killed by accident through for wheeler in north india is greater than accident cause by four wheeler in south india.

ii) $H_{20}: \lambda_2 = \lambda_b$ **against** $H_{21}: \lambda_2 > \lambda_b$

> `poisson.test(c(31,19),c(17,18),alternative = "greater")`

Comparison of Poisson rates

data: c(31, 19) time base: c(17, 18)

count1 = 31, expected count1 = 24.286, p-value = 0.03906

alternative hypothesis: true rate ratio is greater than 1

95 percent confidence interval:

1.033545 Inf

sample estimates:

rate ratio

1.727554

we conduct the test under 0.05% level of significance

we reject Null hypothesis. So $\lambda_2 > \lambda_b$

i.e. the rate of person killed by accident in north india through two/three wheeler is greater than accident in south india cause by two/three wheeler.

iii) $H_{30}: \lambda_3 = \lambda_c$ **against** $H_{31}: \lambda_3 > \lambda_c$


```
> poisson.test(c(32,13),c(17,18),alternative = "greater")
```

Comparison of Poisson rates

```
data:  c(32, 13) time base: c(17, 18)
count1 = 32, expected count1 = 21.857, p-value = 0.001829
alternative hypothesis: true rate ratio is greater than 1
95 percent confidence interval:
 1.464717      Inf
sample estimates:
rate ratio
 2.606335
```

we conduct the test under 0.05% level of significance

we reject Null hypothesis. So $\lambda_3 > \lambda_c$

i.e. the rate of person killed by accident in north India cause by pedestrian and other is greater than compare to accident in south India.

6.CONCLUSION:

From the above analysis we may conclude the following result,

- The rate of person killed in each accident by different types of vehicles(i.e. four/six wheeler, two/three wheeler, pedestrian or non motorized vehicle) is same.
- The rate of death by accident due to different vehicle in north India is greater compare to south India for all types vehicle.
- The cause of accident due to whether over speeding or drunken driving/driving wrong side, the rate of death all cases is equal.
- The rate of death by accident due to different cause (i.e. over speeding,drunken driving/driving wrong side) in north India is greater compare to south India.
- The accident whether happened in sunny, rainy or foggy whether,the death in each accident does not influence by different weather. That means the rate of death in different weather is equal.
- The rate of death by accident due to different weather(i.e. sunny, foggy and rainy) in north India is greater compare to south India.

7.Reference:

1. *Fundamental of Statistics,Vol-1, Vol-2,A.M Gun , M.K Gupta , B. Dasgupta.*
2. *Fundamental of Mathematical Statistics, Gupta & Kapoor.*
3. *Probability Distribution Theory and Statistical inference, K C Bhuyan.*