Behaviour of the urban traffic of the city of Sao Paulo in Brazil

The database was created with records of the urban traffic of the city of Sao Paulo in Brazil from **December 14, 2009 to December 18, 2009** (From Monday to Friday). Registered from 7:00 to 20:00 every 30 minutes.

Dataset – [urban accidents.xlsx](https://docs.google.com/spreadsheets/d/1GM2Yw8OULzBoKY0lkb-8YFkZVvkpa24H/edit?usp=sharing&ouid=100097666359091067095&rtpof=true&sd=true)

Probabilities:

1. Given a certain hour, what is the possibility of the occurrence of an accident?
2. Is slowness of traffic related to the number of accidents occurring? If so, what is the correlation?
3. Taking X = no. of accidents that occur in a half an hour interval,

with X = {0,1,2,3,4,5,6,7,8,10,15}

Given X = no. of accidents

P(X=0) = 35 / 135

P(X=1) = 17 / 135

P(X=2) = 29 / 135  
P(X=3) = 18 / 135

P(X=4) = 14 / 135

P(X=5) = 6 / 135

P(X=6) = 7 / 135

P(X=7) = 5 / 135

P(X=8) = 1 / 135

P(X=10) = 1 / 135

P(X=15) = 2 / 135

Prediction of Probability Distribution: **Poisson distribution**

| **λ (Mean number of accidents)** | 2.518518519 |
| --- | --- |
|  |  |
| **k (number of successes)** |  |
| 0 | 0.080578895 |
| 1 | 0.202939438 |
| 2 | 0.255553366 |
| 3 | 0.214538629 |
| 4 | 0.135079877 |
| 5 | 0.068040234 |
| 6 | 0.028560098 |
| 7 | 0.010275591 |
| 8 | 0.003234908 |
| 9 | 0.000905242 |
| 10 | 0.000227987 |
| 11 | 5.2199E-05 |
| 12 | 1.09553E-05 |
| 13 | 2.1224E-06 |
| 14 | 3.81808E-07 |
| 15 | 6.4106E-08 |

**Graph based on Data Values**

| **k (number of successes)** | **P(X=k)** |
| --- | --- |
| 0 | 0.2592592 |
| 1 | 0.1259259 |
| 2 | 0.2148148 |
| 3 | 0.1333333 |
| 4 | 0.1037037 |
| 5 | 0.0444444 |
| 6 | 0.0518518 |
| 7 | 0.0370370 |
| 8 | 0.0074074 |
| 9 | 0 |
| 10 | 0.0074074 |
| 11 | 0 |
| 12 | 0 |
| 13 | 0 |
| 14 | 0 |
| 15 | 0.0074074 |
|  |  |

Conclusion – The Data doesn’t fit the Poisson Distribution perfectly.

The rapid fall of the probability values indicate that the data could possibly be better modelled with the exponential distribution.

**Checking for Correlation Between the number of accidents occurring and Slowness in Traffic**

| **Hour (Coded)** | **No. Of Accidents over 5 days** | **Slowness in Traffic (%)** |
| --- | --- | --- |
| 1 | 18 | 3.96 |
| 2 | 3 | 5.34 |
| 3 | 4 | 6.64 |
| 4 | 9 | 7.32 |
| 5 | 11 | 9.02 |
| 6 | 16 | 8.88 |
| 7 | 10 | 8.02 |
| 8 | 10 | 7.86 |
| 9 | 12 | 8.18 |
| 10 | 10 | 8.52 |
| 11 | 7 | 8.4 |
| 12 | 6 | 7.7 |
| 13 | 14 | 8.52 |
| 14 | 19 | 7.88 |
| 15 | 18 | 8.34 |
| 16 | 8 | 9.01 |
| 17 | 11 | 9.08 |
| 18 | 21 | 8.7 |
| 19 | 14 | 9.88 |
| 20 | 15 | 11.42 |
| 21 | 7 | 12.66 |
| 22 | 10 | 14.14 |
| 23 | 8 | 15.32 |
| 24 | 14 | 16.62 |
| 25 | 25 | 17.76 |
| 26 | 22 | 17.86 |
| 27 | 18 | 14.38 |

The two variables seem weakly positively correlated.