ANALYSIS OF COMPLAINTS

REGISTERED AT PUNE MUNICIPAL CORPORATION

ABOUT THE TOPIC

‘Complaints being registered at any office’ is a very common phenomenon, though not random. Studying the frequency of complaints, their types, their period of occurrence, etc. surely helps the employers to bring up new policies and proves to be a great step towards the efficient and improved work and here, the Statistics enters the day-to-day life.

Alike every Government office, Pune Municipal Corporation is bound to publish non-sensitive data, and has OPEN DATA PORTAL on its web-sites. I retrieved the data of the complaints to study them and to try and find some key-points to be noted about them, whose further advanced analysis will help the organisation to apply improved method of operandi.

COLLECTION OF DATA

The data that I retrieved was secondary data of the format

Sr. No.	Token No.	Complaint Type	Complaint Subtype	Registration Date	Due Date	Officer	Stage

where, complaint type was varying such as ‘Water supply, Garbage (solid waste management), Health, Garden, etc.’ and subtypes were description of the problem.

The data was non-verbal and needed processing for conversion into numbers

RANDOM VARIABLE

Genus of the random variable concerned with my strategy of analysis could be, say Y

Y : Number of complaints registered on a given day.

Several types of complaints led to several different variables that may or may not be independent.

The variable I considered is

X : Number of complaints registered about Garbage (solid waste management) problem on a given day.

Generally, for the random variable having support from zero to infinity is assumed to follow Poisson distribution when it also satisfy the assumptions like.

- The occurrence of one event does not affect the probability that a second event will occur. That is, events occur independently.
- The rate at which events occur is constant. The rate cannot be higher in some intervals and lower in other intervals.

The Poisson Probability distribution expresses the probability of a given number of events occurring in a fixed interval of time and/or space.

TOOLS USED

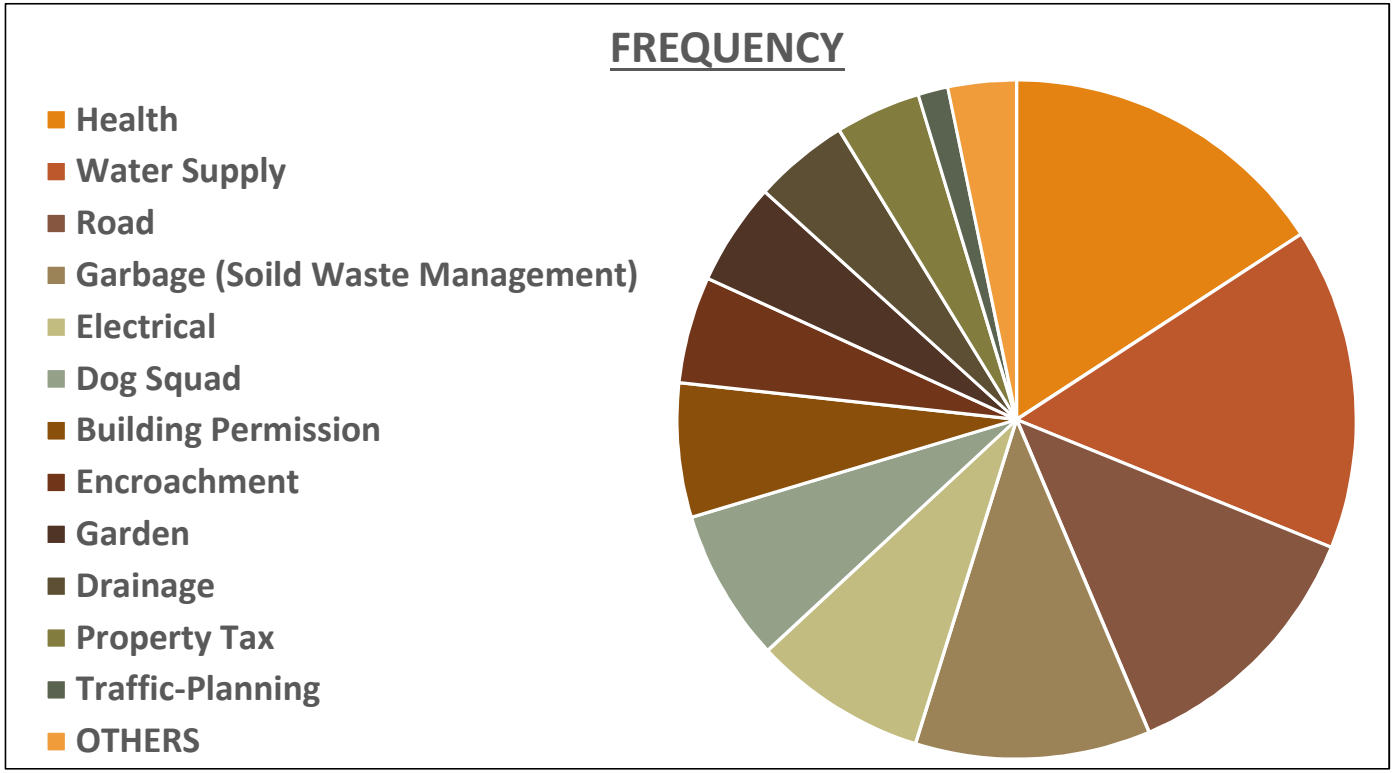
My analysis includes

- Exploratory Data Analysis
- Fitting of distribution using MS Excel
- Test of Goodness of Fit

EXPLORATORY DATA ANALYSIS

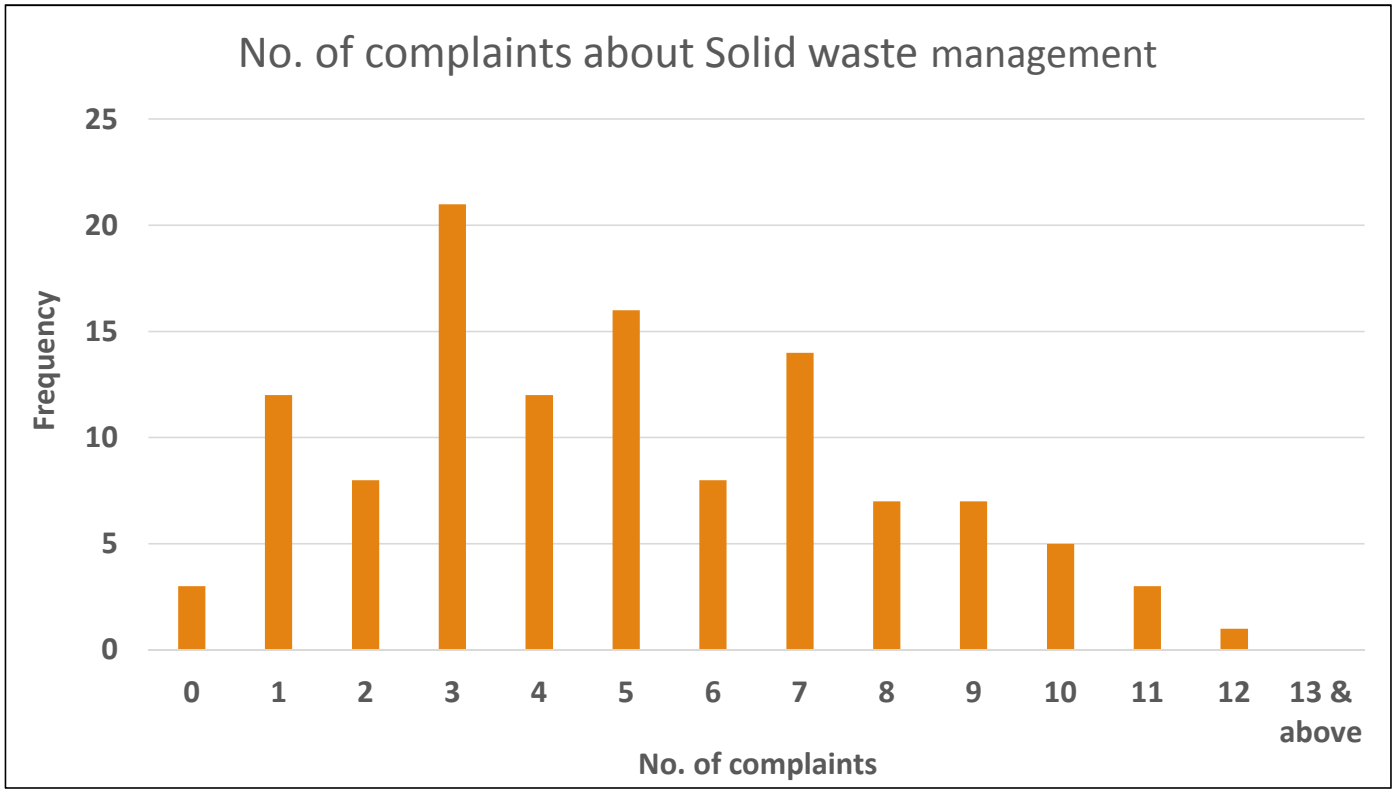
- The representation of the total number of complaints of different types registered

(within a month)



- The representation of the number of complaints of garbage

(date-wise)



FITTING OF THE DISTRIBUTION

Due to the nature of the concerned random variable, the distribution to be fitted is Poisson Distribution.

Let, X : Number of complaints registered about Garbage (solid waste management) problem on a given day.

Let, X follows Poisson distribution with parameter ‘m’

PMF of X is

$$P[X=x] = \frac{(e^{-m}m^{x})}{m!} \quad ; x = 0, 1, 2, \ldots$$

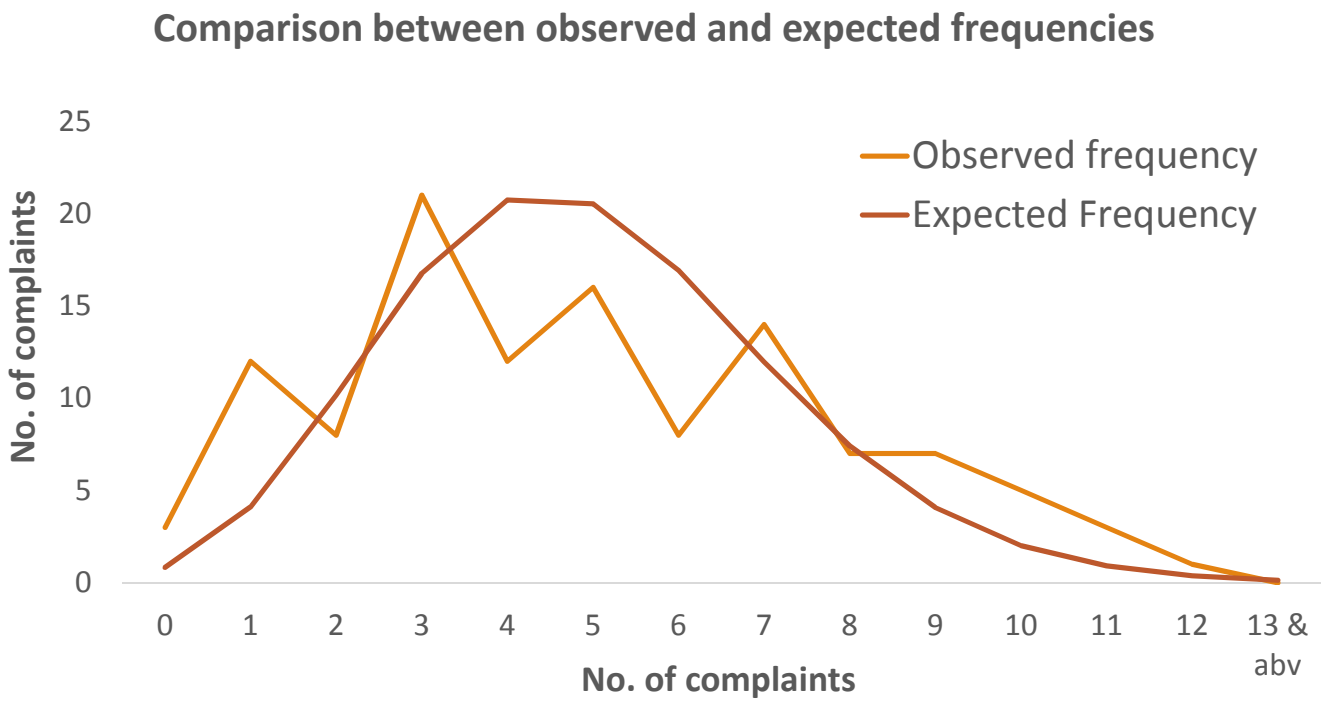
$$= 0 \quad ; \text{otherwise}$$

Frequency distribution of X is

X	Frequency (o_i)	Expected Frequency (e_i)
0	3	0.829881
1	12	4.106832
2	8	10.16174
3	21	16.76247
4	12	20.73811
5	16	20.52533
6	8	16.92895
7	14	11.96804
8	7	7.403282
9	7	4.070736
10	5	2.014485
11	3	0.90628
12	1	0.373742
13 and above	0	0.142272

- Calculated m = 4.9487
- P[X=0] = 0.007093
- Recurrence Relation is
  $P[X=(x+1)] = \{m*P[X=x]\} / (x+1)$
- Expected Frequency is
  $e_i = P[X=x_i]*N$
  where, $N = \Sigma f_i$

- The representation of the observed frequencies and expected frequencies



TEST OF GOODNESS OF FIT

The χ^2 Test of Goodness of fit is a very intuitive name itself. The test is used to check whether the fit of our distribution is good or not.

The test involves χ^2 distribution to determine the test stastistic.

For the test, we define

H_0 : The fit is good for given distribution

Under H_0 , the test statistics is

$$\chi^2_{computed} = \Sigma_i [(o_i - e_i)^2 / e_i]$$

where, o_i = observed frequency foe i^{th} class

e_i = expected frequency for i^{th} class

At level of significance 0.01

The decision rule is

Reject H_0 if,

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

where, k = No. of classes, here k=8 (after pooling the classes)

p = No. of parameters estimated, here p=0

α = Level of significance, here α=0.01

Otherwise, do not reject H_0

For our test of goodness of fit, $\chi^2_{computed} = 24.63691$

$\chi^2_{k-p-1, \alpha} = \chi^2_{8-0-1, 0.01} = 18.475$

Hence, we reject H_0 .

Frequency distribution of X is

k	X	Frequency (o_i)	Expected Frequency (e_i)
1	0	3	0.829881
2	1	12	4.106832
3	2	8	10.16174
4	3	21	16.76247
5	4	12	20.73811
6	5	16	20.52533
7	6	8	16.92895
8	7	14	11.96804
9	8	7	7.403282
10	9	7	4.070736
11	10	5	2.014485
12	11	3	0.90628
13	12	1	0.373742
14	13 and above	0	0.142272

- Class no. (k) 1, 2 and 3 are pooled together as one
- Similarly Class no. (k) 10, 11, 12, 13 & 14 are pooled together as one

As, we reject H_0

The conclusion is, that the fit of the Poisson Distribution for given data is not good

i.e. the random variable may not follow the Poisson Distribution.

This tells us that, the variable may follow any other distribution alike Negative Binomial Distribution that has the same kind of support (from 0 to infinity).

If the H_0 would be accepted, it would be easier to predict the occurrence of the complaints or the probabilities when needed, also, the characteristics of the distribution would highlight the points to consider for their management.

DIFFICULTIES when we deal with the data

- Though the data obtained is authentic, it is secondary type of data, so the errors occurred in collection of data will have no trace to correct them.
- The basic trials were assumed to be Bernoulli trials and their independence had to be assumed as well, the result may not work if they are not.
- The data points involved were very less for such kind of techniques. Large data always gives reliable results, unlike this situation.

MORE TOOLS TO BE USED

- Time series analysis can be done to study the pattern of occurrence and the effect of seasons.
- Method of finding Multiple correlation can be applied to study correlation between different factors.
- Application of Queuing model will be useful to know whether the complaints are being dealt in appropriate time.

Study of these things together will give the real status of the administration and will help to involve changes for betterment of it if needed.

THANK YOU !