RAINFALL DATA ANALYSIS USING BETA DISTRIBUTION

Jeeva .K.G, Ashvin.R, Thejasvin.B.M,Adnan .A, Samuel Naveen .V,Mrs.M.Srividhya

Abstract-Analysation of Rainfall data to check whether it fits to Beta distribution or not. The Rainfall data set taken into account can provide various knowledge depending on users wish. We were able to calculate average monthly rainfall for each year and also the proportion of months in each year where rainfall is above monthly average.The mean and variance of the distribution is calculated along with the estimates of two parameters. It is found out that the Rainfall data obtained fits to the Beta distribution perfectly.

Keywords: Beta distribution, proportion, Mean, Variance and Parameters.

BETA DISTRIBUTION AND APPLICATION

INTRODUCTION:

Rainfall data has been most favourable data. In this case it was simple ,which was to analyse the rainfall data from 1901-2002 for three cities in Tamil Nadu and check whether the beta distribution fits on the data perfectly or not. To do this the attributes like mean, variance, average, proportion and parameters are need to be calculated. For the year 2020 the average rainfall has been interpreted for three cities using linear regression formula. Rainfall data was studied in this paper and the result is to show that it fits beta distribution perfectly.

REASEARCH METHODOLOGY:

The rainfall datasets of Ariyalur, Chennai, Coimbatore from 1901-2002 were taken into account. For each dataset monthly rainfall data was given, the average for each year was calculated. For each year number of months having rainfall more than monthly average is taken into account (r) and proportion is calculated with the formula(r/12). Mean and variance is calculated respectively for the proportions of each year. Beta distribution is given by the formula [(x^(a-1).(1-x)^b-1)/B(a,b)].

$$Y = a + bX$$

$$b = \frac{N\sum XY - (\sum X)(\sum Y)}{N\sum X^2 - (\sum X)^2} \qquad a = \frac{\sum Y - b\sum X}{N}$$

Where,

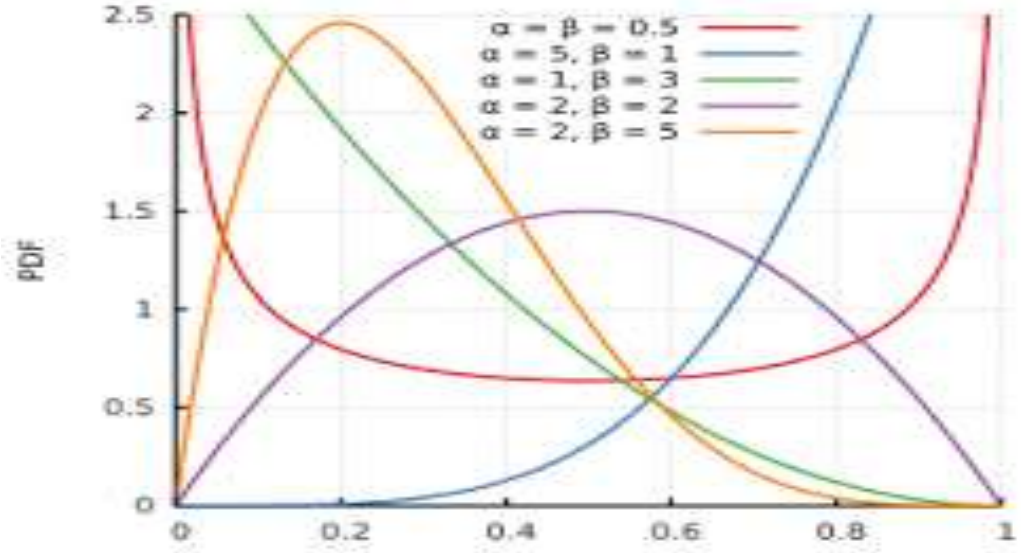
N = number of observations, or years

X = a year index (decade)

Y = population size for given census years

$$
\begin{aligned}
f(x;\alpha,\beta) &= \text{constant} \cdot x^{\alpha-1}(1-x)^{\beta-1} \\
&= \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\,du} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \\
&= \frac{1}{\mathrm{B}(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}
\end{aligned}
$$

$$
r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[\, n\Sigma x^2 - (\Sigma x)^2 \,][\, n\Sigma y^2 - (\Sigma y)^2 \,]}}
$$



DATA ANALYSIS AND INTERPRETATION:

1.ARIYALUR:

---------- TAMIL NADU ---------

---------- ARIYALUR ----------

---------- ESTIMATING THE VALUES OF TWO PARAMETERS OF BETA DISTRIBUTION-

---

THE SAMPLE MEAN IS 72.818353

THE SAMPLE VARIANCE IS 197.379014

ESTIMATED VALUE OF PARAMETER ALPHA IS 2002.191234

ESTIMATED VALUE OF PARAMETER BETA IS 1974.695539

MEAN 0.379902

VARIANCE 0.006614

BETA 1

2.CHENNAI:

---------- TAMIL NADU ---------

---------- CHENNAI ----------

---------- ESTIMATING THE VALUES OF TWO PARAMETERS OF BETA DISTRIBUTION-

---

THE SAMPLE MEAN IS 96.646771

THE SAMPLE VARIANCE IS 499.116037

ESTIMATED VALUE OF PARAMETER ALPHA IS 1886.607446

ESTIMATED VALUE OF PARAMETER BETA IS 1867.086799

MEAN 0.360294

VARIANCE 0.007375

BETA 1

3. COIMBATORE:

---------- TAMIL NADU ---------

---------- COIMBATORE ----------

---------- ESTIMATING THE VALUES OF TWO PARAMETERS OF BETA DISTRIBUTION-

---

THE SAMPLE MEAN IS 186.590815

THE SAMPLE VARIANCE IS 802.678768

ESTIMATED VALUE OF PARAMETER ALPHA IS 8236.578746

ESTIMATED VALUE OF PARAMETER BETA IS 8192.436280

MEAN 0.406046

VARIANCE 0.006219

BETA 1

DATA ANALYSIS AND INTERPRETATION:

Table 1 : ***Actual Data***

YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
1992	3.753	0.14	0.25	26.762	165.914	842.085	700.08	259.483	143.349	286.904	180.041	8.171
1993	0.283	1.39	41.263	57.459	135.939	643.365	526.548	220.347	138.868	308.645	237.658	130.719
1995	5.238	18.952	16.224	64.423	154.356	296.606	595.182	241.309	89.741	323.684	84.68	16.552
1996	7.195	5.808	4.957	183.919	107.682	864.608	398.111	504.735	216.365	284.051	36.378	46.854
1997	12.696	0.136	44.073	88.417	111.134	424.82	361.341	342.961	329.305	380.779	357.168	49.834
1998	1.401	2.368	1.478	132.205	222.878	351.56	815.367	662.227	301.481	349.185	176.69	94.615
1999	3.485	40.691	16.919	127.221	316.837	431.248	459.03	192.056	96.73	351.73	133.476	9.178
2000	1.203	28.567	9.434	132.954	55.513	412.747	370.297	694.243	197.734	338.126	65.536	45.241
2001	4.69	8.357	3.67	246.92	141.457	330.071	607.994	296.327	327.868	232.949	92.123	15.862
2002	0.231	4.669	12.128	33.293	279.139	613.858	157.517	156.434	68.629	365.121	139.152	7.55

Table 2:***MAXIMUM AND MINIMUM RAINFALL AVERAGES OF THREE CITIES***

CITY	MINIMUM RAINFALL	YEAR OF MIN RAINFALL	MAXIMUM RAINFALL	YEAR OF MAX RAINFALL
ARIYALUR	45.00583	2002	112.2966	1988
CHENNAI	50.9577	1905	166.408	1904
COIMBATORE	135.0933	1982	259.2879	1998

Table 3:***TOTAL AVERAGES OF ALL CITIES FROM 1901-2002***

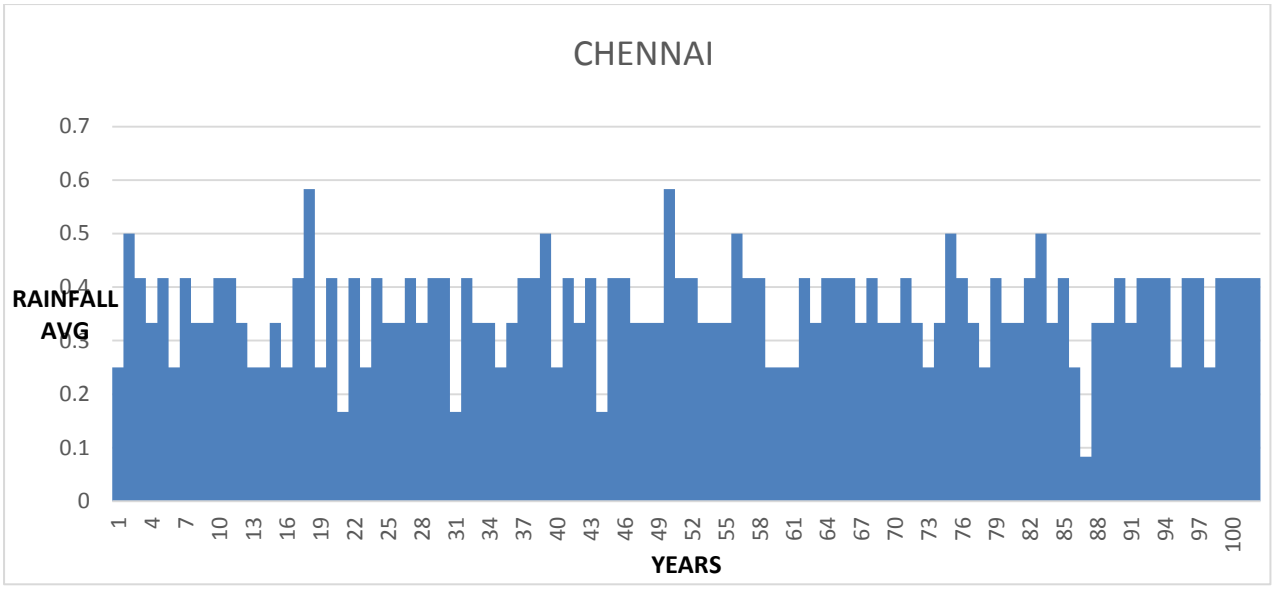
COIMBATORE	186.5908
CHENNAI	96.6468
ARIYALUR	72.81835

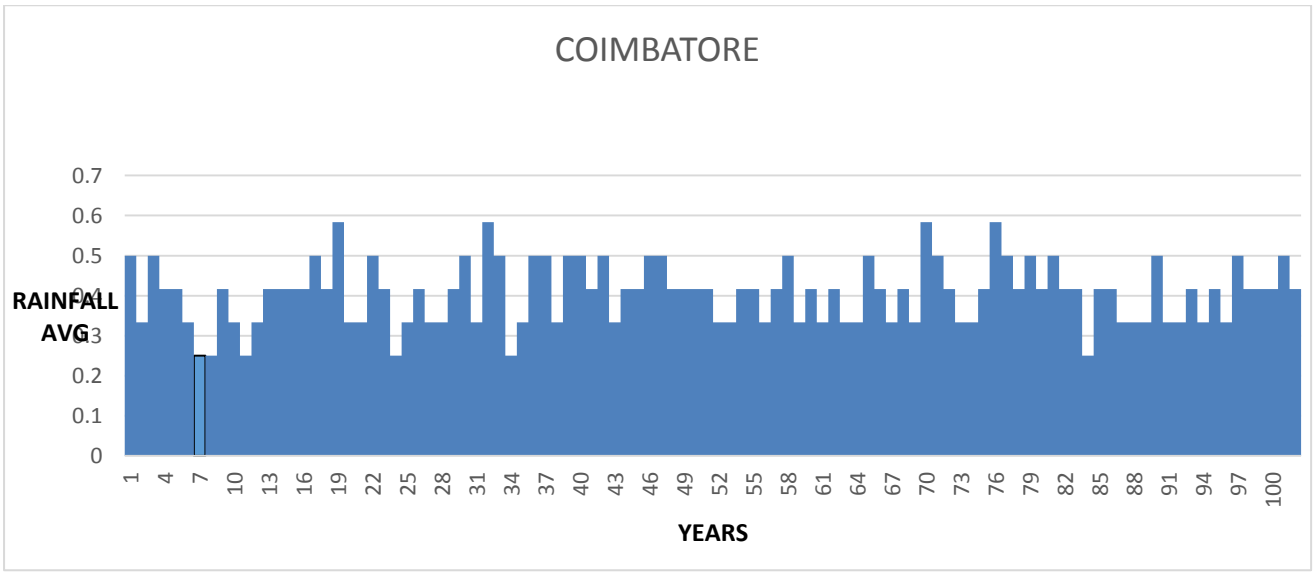
Table 4:***CITIES AND THEIR YEARS WITH HIGHEST & LOWEST PROPORTION***

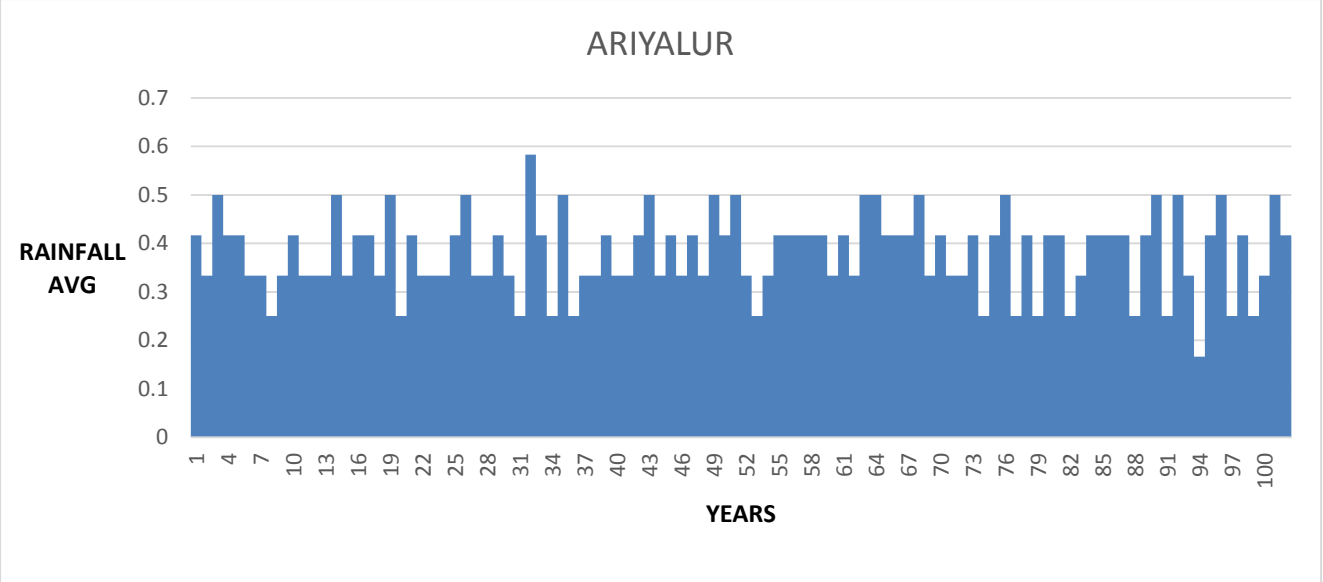
CITIES	HIGHEST PROPORTION	LOWEST PROPORTION
COIMBATORE	1919,1932,1970,1976	1907,1908,1911,1924,1934,1984
CHENNAI	1918,1950	1987
ARIYALUR	1932	1997

Table 5:***CITIES AND THEIR PREDICTION OF AVERAGE RAINFALL IN THE YEAR 2020***

COIMBATORE	187.1714
CHENNAI	100.731
ARIYALUR	74.77807


CHENNAI
RAINFALL AVG
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61 64 67 70 73 76 79 82 85 88 91 94 97 100
YEARS


COIMBATORE
RAINFALL AVG
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61 64 67 70 73 76 79 82 85 88 91 94 97 100
YEARS



CONCLUSION:

The main purpose of this analysation is to fit beta distribution to the rainfall data and the beta distribution ranges from 0 to 1. Since the data analysis resulted in 1 for all the three cities ,it is concluded that the beta distribution fits perfectly to the data and 100% correct.