International Journal of Scientific Research in Mathematical and Statistical Sciences **Research Paper**
Vol.6, Issue.2, pp.228-232, April (2019) **E-ISSN:** 2348-4519
DOI: https://doi.org/10.26438/ijsrmss/v6i2.228232

Probability Analysis of the Rainfall of Dharmavaram Mandal of Anantapuramu District-It's Best Fit

B. Prathap Chowdary[1*], K. Pusphanjali[2], M. C. Chandra Prasad[3]

[1,2,3]Dept. of Statistics, Sri Krishanadevaraya University, Ananthapuramu-515003

[*]*Corresponding Author: b.p.chowdarysas@gmail.com Tel.: 9533460531*

Available online at: www.isroset.org

Received: 18/Apr/2019, Accepted: 26/Apr/2019, Online: 30/Apr/2019

Abstract: This paper studies the probability analysis and estimation of Annual Maximum Monthly Rainfall (AMMR) for different return periods which is very useful for crop planning. We fit various distributions namely NORMAL, LOG NORMAL, LOG PEARSON and GUMBEL for the rainfall data of Dharmavaram Rain Gauge Station (RGS) in Anantapuramu district. Using goodness of fit we determine the best fit probability distribution for predicting the AMMR for different return periods. Based on the obtained numerical results, conclusions are drawn.

Keywords*— *AMMR, Rainfall, Probability distribustions, Goodness of fit, RGS

I. INTRODUCTION

Anantapuramu district lies between 13'- 40' and 15'-15' Northern Latitude and 76'-50' and 78'-30' Eastern Longitude. Dharmavaram is located at 14.43°N 77.72°E. It has an average elevation of 345 meters (1131 feet). It is the hub for pure silk sarees. The economy of the town is dependent on the weaving industry. Farmers depend on rain water due to lack of water resources. A major crop in this area is ground nut.

Rainfall is one of the most important natural input resources for crop production and its occurrence and distribution is unpredictable. Rainfall in this region is sometimes above the normal and sometimes below the normal. (Athi Vrusti and Anaa Vrusti). It receives excess at undesirable times. Hence prediction or estimation of rainfall is the most important factor for crop planning and management of water resources and its applications.

Many researchers worked on these lines for the analysis of rainfall, crop planning, water management and hydrological systems *etc.*

Chow[1] (1964) suggested that rainfall analysis by theoretical probability distributions can be done by using frequency factor 'K_T' based on some statistical parameters.

Bhakar [2] (2008) proposed the scientific prediction of rains and crop planning. Since it was done analytically, it may prove to be a significant tool in the hands of farmers for better economic returns.

Rajendra Subhudhi [3] (2007) found that the normal distribution is the best fit for predicting the annual maximum daily rainfall of Chakapada block of Khadhamal district in Orissa.

M. T. Amin[4] (2016) concluded that the Log Pearson type-III is the best fit probability distribution for the Annual maximum rainfall based on a 24-hour duration at six rainfall-gauging stations in northern Pakistan.

In the present section, the importance of the *Rainfall Analysis* and a brief literature review is presented. In Section-2, we provide the definitions of the probability distributions under study. Data collection and its analysis is provided under section-3. In section-4, we test the goodness of fit for the different distributions under study. Numerical results and discussions are provided under section-5. Finally, In Section-6, we draw conclusions based on the results obtained in the section-5.

II. PROBABILITY DISTRIBUTIONS UNDER STUDY

The probability distribution functions are most commonly used to estimate the rainfall frequency is Normal, Gumbel, Log Pearson type-III, Log Normal distributions.

Normal Distribution:

The normal distribution is the most useful continuous distribution of all the distributions. The probability density function (PDF) and cumulative distribution function (CDF) of the normal distribution are respectively:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right); \quad -\infty < x < \infty \text{ .... (2.1)}$$

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{2}\left[1 + erf\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right] \text{.................. (2.2)}$$

Where 'μ' is the location parameter, 'σ' is the scale parameter and 'Φ' is the Laplace Integral.

Log Normal distribution:

The log-normal distribution is a distribution of random variables with a normally distributed logarithm. The log-normal distribution model includes a random variable Y, and Log(Y) is normally distributed. The probability density function (PDF) and cumulative distribution function (CDF) of the log-normal distribution are respectively

$$f(x) = \frac{exp\left[-\frac{1}{2}\left(\frac{\ln(x-\gamma)-\mu}{\sigma}\right)^2\right]}{(x-\gamma)\sigma\sqrt{2\pi}}; \quad 0 < x < \infty \text{ .... (2.3)}$$

$$F(x) = \Phi\left(\frac{\ln(x-\gamma)-\mu}{\sigma}\right) = \frac{1}{2}\left[erfc\left\{-\frac{\ln(x-\gamma)-\mu}{\sigma\sqrt{2}}\right\}\right] \text{ .... (2.4)}$$

where 'μ' is the shape parameter, 'σ' is the scale parameter, 'γ' is the location parameter and 'Φ' is the Laplace Integral.

Log-Pearson type-III distribution:

The log-Pearson type-III distribution has been widely and frequently used in hydrology and for hydrologic frequency analyses since the recommendation of this distribution by U.S. Federal agencies. The probability density function (PDF) and cumulative distribution function (CDF) of the Log-Pearson type-III distribution are respectively

$$f(x) = \frac{1}{x/\beta/\Gamma(\alpha)}\left(\frac{\ln(x)-y}{\beta}\right) \exp\left(-\frac{\ln(x)-y}{\beta}\right)^{\alpha-1}; 0 < x < \infty \text{ ....(2.5)}$$

$$F(x) = \frac{\Gamma_{\frac{(\ln(x)-y)}{\beta}}(\alpha)}{\Gamma(\alpha)} \text{ ....(2.6)}$$

where 'α', 'β' and 'γ' are shape, scale and location parameters, respectively.

Gumbel (EV I) distribution

The Gumbel distribution named in honor of Emil Gumbel, and also known as the Extreme Value Type I (EVI) distribution, is a continuous probability distribution. This distribution can be applied to model maximum or minimum values (extreme values) of a random variable. The probability density function (PDF) and cumulative distribution function (CDF) of the Gumbel distribution are respectively:

$$f(x) = \frac{1}{\sigma} \exp\left(-\frac{x-\mu}{\sigma} - exp\left(-\frac{x-\mu}{\sigma}\right)\right); -\infty < x < \infty \text{ .... (2.7)}$$

$$F(x) = exp\left(-exp\left(-\frac{x-\mu}{\sigma}\right)\right) \text{ .... (2.8)}$$

Where 'σ' and 'μ' are the scale and location parameters respectively.

III. DATA COLLECTION AND ANALYSIS

Annual Monthly Maximum Rainfall data of Dharmavaram RGS has been used for the present investigation. Time series rainfall records for the period of 39 years (1980 to 2018) have been collected from the Chief Planning Office, Anantapuramu. The maximum amount of monthly rainfall for each year has taken for the analysis.

The parameters are used generally mean, variance, standard deviation; coefficient of variation and coefficient of skewness were taken as measures of any type of statistical analysis. All these parameters have been used to describe the variability of rainfall in the present study.

Return period was calculated by Weibull's plotting position formula given by Chow (1964). Expected AMMR are obtained for the return periods of 1.01, 1.05, 1.11, 1.25, 2, 4, 5, 10, 20, 40, 100 and 200 years. These values of AMMR can be obtained statistically through the Chow's general frequency formula. The formula expresses the frequency of occurrence of an event in terms of a frequency factor K_T, which depends upon the distribution of particular event investigated. Chow (1951) has shown that many frequency analyses can be reduced to the form

$$X_T = \bar{X}\,(1 + C_v K_T) \text{ .... (3.1)}$$

Where $\bar{X}$ is mean of the monthly maximum rainfall of the data of length N years, C_v is the coefficient of variation and K_T is the frequency factor which depends upon the return period T and the assumed frequency distribution. The expected value of AMMR for the same return periods were computed for determining the best probability distributions.

IV. TESTING OF THE GOODNESS OF FIT

The expected values of AMMR were calculated by well-known probability distributions which are mentioned above in section II. The probability distributions under study were tested by the most commonly used Goodness of fit namely

Chi-Square test, Anderson-Darling test and Kolmogorov-Smirnov test.

Chi-square Test:
Chi-square test is most commonly used to test the goodness of fit of any probability distributions. The test statistic of chi-square test is given by

$$\chi^2 = \sum_{i=1}^{k} \frac{(\mathrm{Oi} - \mathrm{Ei})^2}{\mathrm{Ei}} \ldots\ldots\ldots\ldots\ldots\ldots \quad (4.1)$$

Where, k is the number of class intervals, O_i and E_i are the observed and expected rainfall values in the i^{th} class respectively. The best probability distribution function was determined by comparing chi-square values obtained from each distribution and selecting the function that gives which has least sum of chi-square values will be a good fit.

Anderson-Darling Test:
This test was proposed by Stephens (1974), it is used to test the given data follow a specified distribution or not. The test statistic is defined as

$$A^2 = -N - S \ldots\ldots\ldots\ldots (4.2)$$

Where

$$S = \sum_{i=1}^{n} \frac{(2i-1)}{N} [lnF(Y_I) + \ln(1 - F(Y_{N+1-i}))] \ldots (4.3)$$

F is the cumulative distribution function of the specified distribution. Y_i are the ordered data. The critical values for the AD test are dependent on the specific distribution, It is being tested at α level of significance. Tabulated values are published by Stephens (1979) for the distribution of the test statistic A is greater than the critical value then the null hypothesis H_0 is rejected.

Kolmogorov-Smirnov Test:
The Kolmogorov-Smirnov test used to test whether sample comes from a population with specified theoretical distribution. It is based on the empirical distribution function (ECDF). Given N ordered statistics $Y_1, Y_2 \ldots.. Y_N$, the ECDF is defined as

$$E_N = \frac{n(i)}{N} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (4.4)$$

Where $n(i)$ the number of points is less than Y_i and the Y_i are ordered from smallest to largest. The most important feature of this test is the distribution of the K-S test statistic itself does not depend upon the underlying cumulative distribution function being tested.

The test statistic is defined by

$$D = \max_{1\le i\le N}\left(F(Y_i) - \frac{i-1}{N}, \frac{i}{N} - F(Y_i)\right) \ldots\ldots\ldots\ldots (4.5)$$

Where F is the theoretical cumulative distribution function of the distribution being tested which must be a continuous distribution at α level of significance. If the statistic value D is greater than the critical value the null hypothesis is rejected.

V. NUMERICAL RESULTS AND DISCUSIONS

The distribution of monthly rainfall observed during the period 1980 to 2018 is shown in the Figure-5.1.

Distribution of Monthly Rainfall

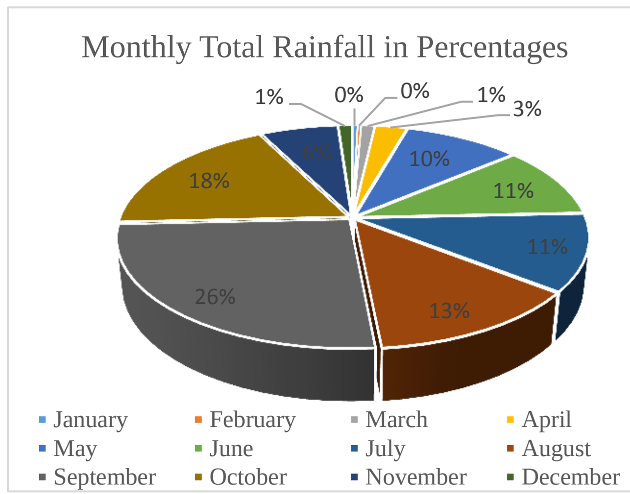

Figure 5.1

The four probability distributions (Normal, Log-Normal, Log-Pearson type-III and Gumbel) were used in this study. The Parameters of probability distributions were calculated using the method of moments and are given in Table 5.1.

Table 5.1
Estimated Parameters of Probability distributions

Distributions	**Estimated Parameters**	**Dharmavaram Gauging station**
Normal	sigma(σ)	88.842
	mu(μ)	199.69
Log-Normal	sigma(σ)	0.22197
	mu(μ)	5.9347
	Gamma(γ)	-187.67
Log-Pearson type III	Alfa(α)	3.9041
	Beta(β)	-0.25602
	Gamma(γ)	6.1897
Gumbel	sigma(σ)	69.27
	mu(μ)	159.71

Statistical parameters used to calculate the AMMR for different return periods and its summary of the statistics is presented in the following Table-5.2.

Table 5.2
Summary of Statistics of Dharmavaram RGS

	Statistical Parameters	
1	Mean	199.69
2	Standard Error	14.23
3	Median	188.20
4	Standard Deviation	88.84
5	Sample Variance	7892.84
6	Kurtosis	2.09
7	Skewness	0.88
8	Minimum	35.50
9	Maximum	500.00
10	Count	39.00
11	Coefficient of variation	0.39
12	Coefficient of skewness	0.38

The obtained mathematical expression for the different distributions under study is presented in the following equations

1. Normal distribution :

$$X_T = 199.7 * (1 + 0.44 * K_T)\text{………… (I)}$$

2. Log Pearson distribution:

$$\log(X_T) = 199.7 + (K_T * 88.84)\ \text{………….. (II)}$$

3. Gumbel distribution :

$$X_T = 199.7 * (1 + 0.44 * K_T)\text{……. (III)}$$

4. Log Normal distribution :

$$Y_T = 5.18 * \left(1 + (K_T * 0.09)\right)\text{………….. (IV)}$$

The expected AMMR for the different return periods using the above mathematical expressions are shown in the following Table 5.3.

Table 5.3
Expected AMMR for the Different Return Periods

S.no	observed values	Return Period	Normal	Log Pearson	Gumbel	Log Normal
1	35.5	1.02	39.82	53.24	69.35	72.1
2	54.4	1.05	61.46	70.28	83.76	81.55
3	94.1	1.11	89.43	93.21	101.98	95.63
4	132.6	1.25	126.06	124.81	126.78	126.78
5	188.2	2	199.67	193.94	185.11	179.14
6	263.2	4	259.55	255.96	246	251.93
7	278.2	5	274.4	272.18	263.59	274.17
8	314.1	10	313.51	316.49	315.55	342.55
9	326.2	20	345.79	354.88	365.39	411.68
10	500	40	373.79	389.52	414.28	482.82
11		100	406.33	431.4	478.25	581.11
12		200	428.49	460.92	526.42	659.25

The four probability distributions were subjected to three goodness of fit tests to determine the best-fitting probability model for rainfall gauging station. The goodness of fit tests was ranked from one (best-fit) to four (least fit) for all probability distributions are shown in the following Table 5.4.

Table 5.4
Scores of Goodness of Fit for Various Distribution Models

Model	Ks-Test	$\chi 2$ −Test	AD Test	Total
Normal	4	2	3	9
Log-Pearson	3	1	4	8
Gumbel	2	4	2	8
Log-Normal	1	3	1	5

Selection of best fit probability distribution is based on total score from all the goodness of fit tests.

VI. CONCLUSIONS

We draw the following conclusions from the obtained results discussed in Section V.

1. From the Table5.4 Normal distribution provides best fit for the Dharmavaram RGS and the mathematical expression of best fit distribution is

$$X_T = 199.7 * (1 + 0.44 * K_T)$$

2. The best fitted Normal model is obtained as

$$f(x, \hat{\mu}, \hat{\sigma}) = \frac{1}{88.842\sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{x-199.69}{88.842}\right)^2\right)$$ -□ < x< □

Where $\hat{\mu}$ *and* $\hat{\sigma}$ moment estimators of μ, σ.

3. The curve of the best fit distribution of Dharmavaram rain gauge station shown in Figure-6.1.

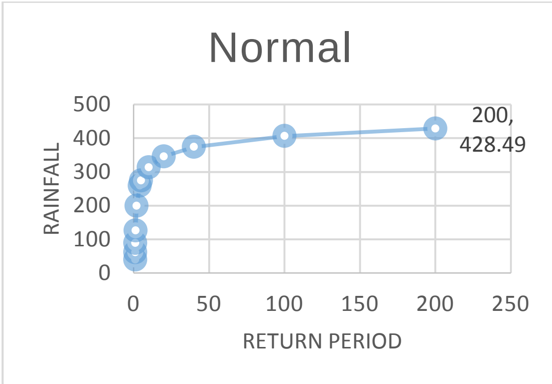

Figure 6.1

4. From the Figure-5.1. It shows that 26% and 18% of observed rainfall in September and August. These results suggest that the Dharmavaram Rain gauge station receives maximum rainfall during August and September. It is the best period for the crop planning of Dharmavaram Mandal.

REFERENCES

[1] Chow, V.T., 1964.handbook of applied hydrology.

[2] Bhakar S. R., Iqbal M., Devanda M., Chhajed N., Bansal A. K., Probability analysis of rainfall at Kota, Indian J. Agri. Res, 2008, 42, 201 -206.

[3] Subudhi R., Probability analysis for prediction of annual maximum daily rainfall of Chakapada block of Kandhamal district in Orissa, Indian J. Soil Conser, 2007, 35, 84-85.

[4] M. T. Amin*, M. Rizwan, A. A. Alazba, A best-fit probability distribution for the estimation of rainfall in northern regions of Pakistan, Open Life Sci. 2016; 11: 432–440.

[5] Kazi S. A., Khan M. L., Variability of rainfall and its bearing on agriculture in the arid and semi-arid zones of West Pakistan, Pak. Geographic Rev., 1951, 6 (1), 40-63.

[6] Manikandan M., Thiyagarajan G., Vijayakumar G., Probability analysis for estimating annual One day maximum rainfall in Tamil Nadu Agricultural University, Mad. Agri. J., 2011, 98 (1-3), 69-73.

[7] Upadhaya A., Singh S. R., Estimation of consecutive day's maximum rainfall by various methods and their comparison, Indian J. Soil Conserv., 1998, 26, 193-201.

[8] Bhakar, S.R., Bansal, A.N, Chhajed, N. Purohit, R.C. 2006. Frequency analysis of consecutive day maximum rainfall at Banswara, Rajasthan India. ARPN J. Eng. Appl. Sci., 1: 64-67.

[9] Agarwal, M. C., V. S. Katiyar and Ram Babu (1988). Probability analysis of annual maximum daily rainfall of U. P., Himalaya. Indian Journal of Soil Conservation, 16(1): 35-42.

[10] Gupta, S. C. and V. K. Kapoor (2002). Fundamental of Mathematical Statistics, Sultan Chand and Sons, New Delhi.

[11] Murray R.S., Larry J. S., Theory and problems of statistics, 3rd Edition, Tata Mc Graw – Hill Publishing Company Limited, New Delhi, India, 2000, 314 – 316.

[12] Chowdhury J. U., Stedinger J. R., Goodness of fit tests for regional generalized extreme value flood distributions, Water Res., 1991, 27 (7), 1765 – 1777.

AUTHERS PROFILE

Mr. B. Prathap Chowdary is pursuing M.Phil. in statistics from Sri Krishnadevaraya University, Anantapuramu, Andhra Pradesh, India. His research topic is "Rainfall analysis of selected rain gauge stations in Anthapuramu district". He has 4 years of experience in teaching Statistics for graduates and CA-CPT students.

Dr. K. Pushpanjali presently working as Professor in the department of Statistics, Sri Krishnadevaraya University. She has 33 years of teaching experience and published 16 research papers in national and international journals. She has guided one M.Phil. and three Ph.D. Scholars. Here area of research interest are Distribution theory, Statistical inference, Reliability and Data envelopments analysis.

M. C. Chandra Prasad is pursuing Ph. D. in Statistics from Sri Krishnadevaraya University, Anantapuramu, Andhra Pradesh, India. His research topic is "CASP – CUSUM Schemes based on Truncated Mixed Weibull Distributions". He has 5 years of experience in teaching Statistics for Graduates and CA-CPT students. He has published two papers in reputed international journals.

