ORIGINAL RESEARCH PAPER

Statistics

FITTING OF NEGATIVE BINOMIAL DISTRIBUTION BY USING RECURRENCE RELATION METHOD BETWEEN RAINFALL AND GROUND WATER LEVELS – A CASE STUDY

KEY WORDS: Rainfall, Ground Water Level, RSS, Validation of the distribution, C.D test.

Raju Sake Academic Consultant, Department of Statistics, Sri Krishnadevaraya University, Anantapuramu (A.P), India.

ABSTRACT

Present paper deals with the application of Distribution Theory to analyze and predict Rainfall (RF) and Ground Water Levels (GWLs) in Anantapuramu district based on the data collected from January 2007 to December 2016. Through with Negative Binomial Distribution by using Recurrence Relation Method for the purpose of analysis the district is divided into five zones. We have estimated the Negative Binomial Distribution by using Recurrence Relation values and compared among them by using the data. Further, validation of the fitted distribution identified the best suitable zone that is Residual Sum of Squares (RSS) value of the zone and forecast on the Rainfall and Ground Water Levels of this district. We also calculate Critical Difference (C.D) test and conclusions are drawn based on the results obtained.

INTRODUCTION:

We have discussed 'Distribution Theory' for different distributions like. Binomial Distribution-Direct and Recurrence Relation Method already we will analyze, now we will fit **Negative Binomial Distribution Recurrence Relation Method in this paper.**

The data is collected on Average RF and Average GWLs are given in my previous published research papers for ready reference [1,2,3,4,5 and 6].

STATISTICAL ANALYSIS

To forecast **Rainfall** and **Ground Water Levels** through Negative Binomial Distribution for different zones we can consider given as follows:

The Probability Mass Function of Negative Binomial Distribution is given by

$$P(X=x) = {}^{x+r-1}_{r-1}c\, p^r q^x$$

(Or)

$$P(X=x) = {}^{x+r-1}_{x}c\, p^r q^x \quad where\ x = 1,2,3,\ldots, \quad \ldots\ldots (2.1)$$

Here r and p are the parameters of Negative Binomial Distribution. To examine these parameters, we will require the Mean ($\bar{x}$) and Variance (σ^2) of the given frequency distribution.

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{N};\ \sigma^2 = \frac{1}{N}\sum f_i x_i^2 - (\bar{x}^2)$$

In Negative Binomial Distribution we know that $\bar{x} = \frac{rq}{p};\ \sigma^2 = \frac{rq}{p^2};\ \frac{\bar{x}}{\sigma^2} = p = \hat{p};\ \hat{q} = 1 - \hat{p}$

$$\bar{x} = \frac{r\hat{q}}{\hat{p}};\ r = \frac{\hat{p}\bar{x}}{\hat{q}}$$

Substitute these $\hat{p}$, $\hat{q}$ in the above Negative Binomial Distribution equation (2.1) we will get the Probability Mass Function of a Negative Binomial Distribution has

$$P(X=x) = {}^{x+r-1}_{x}c\, \hat{p}^r \hat{q}^x \quad where\ x = 1,2,3,\ldots, \quad \ldots\ldots (2.2)$$

To find the Expected Frequencies: To find the Expected Frequencies, we use the following Recurrence Relation formula for Negative Binomial Distribution is given by

$$P(x+1) = \frac{x+r}{x+1}\, q\, p(x) \qquad \ldots\ldots (2.3)$$

The fitted Negative Binomial Distribution by using Recurrence Relation Method for Average RF and Average GWLs:

A: For Average Rainfall

Zone-I

The Probability Mass Function of Negative Binomial Distribution is given by

$$P(X=x) = {}^{x+7-1}_{x}c\, (0.59)^7 (0.41)^x$$

In Recurrence Relation

$$P(x+1) = \frac{x+7}{x+1}\,(0.41)\,p(x)$$

Zone-II

The Probability Mass Function of Negative Binomial Distribution is given by

$$P(X=x) = {}^{x+8-1}_{x}c\, (0.61)^8 (0.39)^x$$

In Recurrence Relation

$$P(x+1) = \frac{x+8}{x+1}\,(0.39)\,p(x)$$

Zone-III

The Probability Mass Function of Negative Binomial Distribution is given by

$$P(X=x) = {}^{x+8-1}_{x}c\, (0.61)^8 (0.39)^x$$

In Recurrence Relation

$$P(x+1) = \frac{x+8}{x+1}\,(0.39)\,p(x)$$

Zone-IV

The Probability Mass Function of Negative Binomial Distribution is given by

$$P(X=x) = {}^{x+8-1}_{x}c\, (0.62)^7 (0.38)^x$$

In Recurrence Relation

$$P(x+1) = \frac{x+8}{x+1}\,(0.38)\,p(x)$$

Zone-V

The Probability Mass Function of Negative Binomial Distribution is given by

$$P(X=x) = {}^{x+7-1}_{x}c\, (0.59)^7 (0.41)^x$$

In Recurrence Relation

$$P(x+1) = \frac{x+7}{x+1}\,(0.41)\,p(x)$$

B: For Average Ground Water Levels

Zone-I

The Probability Mass Function of Negative Binomial Distribution is given by

$$P(X=x) = {}^{x+19-1}_{x}c\, (0.76)^{19} (0.24)^x$$

In Recurrence Relation

$$P(x+1) = \frac{x+19}{x+1}\,(0.24)\,p(x)$$

Zone-II

The Probability Mass Function of Negative Binomial Distribution is given by

$$P(X=x) = {}^{x+11-1}_{x}c\, (0.65)^{11} (0.35)^x$$

In Recurrence Relation

$$P(x+1) = \frac{x+11}{x+1}\,(0.35)\,p(x)$$

Zone-III

The Probability Mass Function of Negative Binomial Distribution is given by

$$P(X=x) = {}^{x+21-1}_{x}c\, (0.77)^{21} (0.23)^x$$

In Recurrence Relation

$$P(x+1) = \frac{x+21}{x+1}\,(0.23)\,p(x)$$

Zone-IV

The Probability Mass Function of Negative Binomial Distribution is given by

$$P(X=x) = {}^{x+14-1}_{x}c\, (0.69)^{14} (0.31)^x$$

In Recurrence Relation

$$P(x+1) = \frac{x+14}{x+1}(0.31)\,p(x)$$

Zone-V

The Probability Mass Function of Negative Binomial Distribution is given by

$$P(X=x) = {}^{x+12-1}_{x}c\,(0.67)^{12}(0.33)^{x}$$

In Recurrence Relation

$$P(x+1) = \frac{x+12}{x+1}(0.33)\,p(x)$$

where $x = 1,2,3\ldots$ Substitute in the above equations we can get the values of $p(1), p(2), p(3), \ldots,$ *multiplying these* $p(1), p(2), p(3), \ldots,$ *values by the* $N = \sum_{i=1}^{n} f_i$ we get the required Expected Frequencies, these are denoted by $f(1), f(2), f(3), \ldots$

VALIDATION OF THE FITTED DISTRIBUTION

Validation of the fitted distribution is necessary to check the suitability of the distribution for the given data this is done by considering X = Years and Y = Average RF or Average GWL and estimated the Average RF (Y) or Average GWL (Y) denoted by ŷ. The estimated Average RF and Average GWLs are given in the following tables.

Table-3.1 Estimated Average RF ŷ For Negative Binomial Distribution By Using Recurrence Relation Method

Year	Zone-I		Zone-II		Zone-III		Zone-IV		Zone-V	
	Actual	Estimates	Actual	Estimates	Actual	Estimates	Actual	Estimates	Actual	Estimates
2007	65.60	31.16	58.20	31.35	67.20	30.99	52.00	31.56	60.50	30.94
2008	53.90	48.97	77.90	57.48	65.20	56.82	61.30	54.11	62.70	48.62
2009	45.40	62.33	50.60	73.15	46.30	72.31	57.10	67.64	38.70	61.88
2010	53.90	62.33	71.50	78.38	70.80	77.48	64.60	72.14	56.30	61.88
2011	39.50	57.88	42.30	73.15	48.90	72.31	31.80	67.64	36.60	57.46
2012	43.20	48.97	43.40	62.70	45.30	61.98	40.50	54.11	41.90	48.62
2013	35.00	35.62	52.30	47.03	47.10	46.49	34.80	40.58	38.10	35.36
2014	31.10	26.71	30.30	36.58	27.10	36.16	37.10	27.05	22.80	26.52
2015	44.10	17.81	62.60	26.13	66.30	25.83	46.00	18.04	54.30	17.68
2016	33.50	13.36	33.40	15.68	32.30	15.50	25.70	13.53	30.10	13.26

Table-3.2 Estimated Average GWL ŷ For Negative Binomial Distribution By Using Ecurrence Relation Method

Year	Zone-I		Zone-II		Zone-III		Zone-IV		Zone-V	
	Actual	Estimates	Actual	Estimates	Actual	Estimates	Actual	Estimates	Actual	Estimates
2007	10.57	2.71	22.58	6.36	14.23	3.10	14.97	2.83	17.03	4.16
2008	9.96	6.77	20.73	12.73	9.27	7.76	10.88	7.07	9.09	8.32
2009	12.17	10.83	17.53	19.09	11.08	13.97	9.58	11.31	10.24	12.48
2010	12.74	14.89	15.02	23.33	12.03	18.62	8.58	15.55	11.79	15.25
2011	12.69	16.25	15.20	25.46	11.48	21.72	8.93	16.96	12.84	16.63
2012	14.98	16.25	20.49	23.33	16.08	21.72	13.76	16.96	13.22	15.25
2013	15.94	13.54	23.03	19.09	18.69	18.62	16.98	15.55	14.30	12.48
2014	15.87	[illegible]	23.40	14.85	21.16	15.52	18.92	12.72	16.30	9.70
2015	14.90	8.12	26.88	10.61	25.80	10.86	19.26	9.90	17.66	6.93
2016	15.57	5.42	27.27	8.49	15.35	7.76	19.51	7.07	16.15	4.16

In the above tables -3.1 and 3.2 for the validation of the distribution, RSS calculated zone wise by considering

$$Residual\ Sum\ of\ Squares\ (RSS) = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \ldots..(3.1)$$

Where y_i or o_i represents actual or observed values and ŷ or ê is the estimated values through fitted distribution is given in tables- 3.1 and 3.2. RSS was calculated and is given in the following table.

Table-3.3 RSS Values For Average RF For Negative Binomial Distribution.

Type of the Distribution	Zone-I	Zone-II	Zone-III	Zone-IV	Zone-V
Negative Binomial Distribution	3055.63	4729.21	4931.28	3171.45	3766.75

Table-3.4 RSS Values For Average GWLs For Negative Binomial Distribution.

Type of the Distribution	Zone-I	Zone-II	Zone-III	Zone-IV	Zone-V
Negative Binomial Distribution	272.81	1217.94	627.23	571.03	507.46

CONCLUSIONS:

By Comparing RSS values for Average RF and Average GWLs through Negative Binomial Distribution by using Recurrence Relation formula under consideration, for RF of zone-I is least and GWLs for zone-I RSS values is least. Next to zone-I, zone-IV has least RSS value in RF and GWLs zone-V is least. Further, the behaviors of RF and GWL through this distribution in different zones are represented in the following Figure-3.1. Similar conclusions can be drawn from the following graphs also.

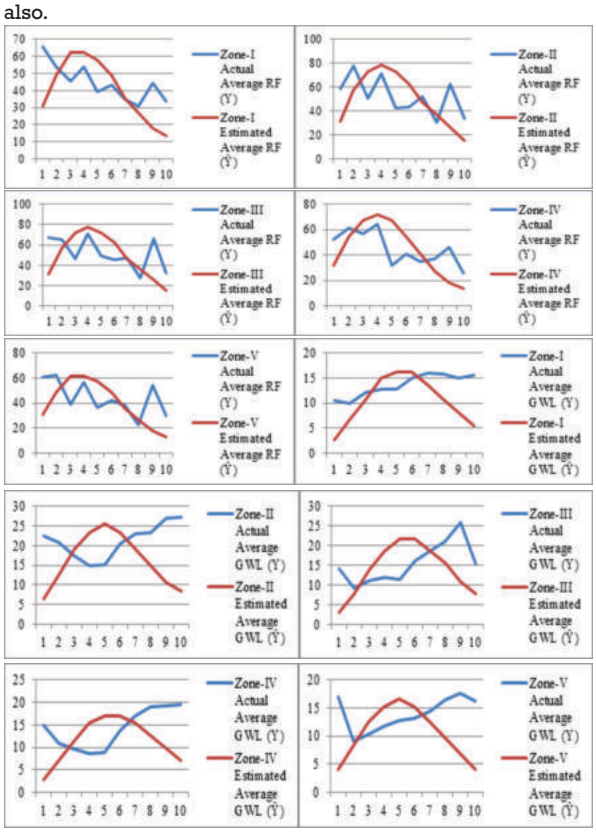

Fig-3.1 Behavior Of RF And GWLs Actual And Estimated Values For Negative Binomial Distribution By Using Recurrence Relation Method In Zone –I, II, III, IV and V

Note: In the above graphs x-axis represents years in the last decade i.e. from 2007 to 2016. On y-axis Average RF measured in Mille Meters or Average GWLs measured in Meters [1, 2, 3, 4, 5 and 6].

FURTHER STATISTICAL ANALYSIS

Now, we proceed to analyze the given estimates in tables-3.1 and 3.2 using ANOVA two-way classification by considering rows as different years and columns as different zones and the following Null Hypothesis are formed and tested [1, 2, 3, 4, 5 and 6].

H_{01}: There is no significant difference between different years of Average RF in Anantapuramu District [1, 2, 3, 4, 5 and 6].
H_{02}: There is no significant difference between Average RF of different zones in Anantapuramu District [1, 2, 3, 4, 5 and 6].
H_{03}: There is no significant difference between different years of Average Ground Water Levels in Anantapuramu District [1, 2, 3, 4, 5 and 6].
H_{04}: There is no significant difference between Average Ground Water Levels of different zones in Anantapuramu District [1, 2, 3, 4, 5 and 6].

Table-4.1 ANOVA Two-way Table for RF

Source of variation	d.f	S.S	M.S.S	F-cal
Rows (years)	9	18410.51	2045.612	269.7763
Columns (Zones)	4	907.7704	226.9426	29.92931
Error	36	272.9744	7.582621	
Total	49	19591.25		

By comparing F-calculated value of Rows (Years) with F-critical value at 5% level of significance we reject the H_{01} i.e. There is a significant difference between different years of Average RF in Anantapuramu District. Similarly by comparing F-calculated value of Columns (Zones) with F-critical value at 5% level of significance we reject the H_{02} i.e. There is a significant difference between Average RF of different zones in Anantapuramu District.

Table-4.2 ANOVA Two-way Table For GWLs

Source of variation	d.f	S.S	M.S.S	F-cal
Rows (years)	9	1299.93	144.4366	67.59983
Columns (Zones)	4	252.3963	63.09908	29.53189
Error	36	76.91911	2.136642	
Total	49	1629.245		

By comparing F-calculated value of Rows (Years) with F-critical value at 5% level of significance we reject the H_{03} i.e. There is a significant difference between different years of Average GWLs in Anantapuramu District. Similarly by comparing F-calculated value of Columns (Zones) with F-critical value at 5% level of significance we reject the H_{04} i.e. There is a significant difference between Average GWLs of different zones in Anantapuramu District.

Since F-cal value related to Rows (Years) in RF is high so there is a necessity for Critical Difference (C.D) Test for sub-grouping various years using the following formula.

C.D. $= \sqrt{2 \times Error M.S.S/m} \times t_{0.01}$ for error d.f. in tables (4.1) and (4.2) ... (4.1)
Where m represents number of replicates in each zone and as well as year.

CRITICAL DIFFERENCE (C.D) TEST: Average RF for Year

Table-5.1 Year Wise Aggregate Average RF For Negative Binomial Distribution Estimates

Year	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Average	31.2	53.2	67.462	70.442	65.688	55.276	41.016	30.604	21.098	14.266

Table 5.2 If We Can Arranged Ascending Order

Year	2016	2015	2014	2007	2013	2008	2012	2011	2009	2010
Average	14.266	21.098	30.604	31.2	41.016	53.2	55.276	65.688	67.462	70.442

S.E $= \sqrt{2 \times Error M.S.S/m}$
$= 1.74$
1% l.o.f C.D $= 2.58 \times 1.74 = 4.49$

2016 2015 2014 2007 2013 2008 2012 2011 2009 2010

Above notation indicates that 2014-2007, 2008-2012, 2011-2009 years Average RF come under one category and 2016, 2015, 2013, 2010 year Average RF different category because there is no Significant Difference in average RF.

REFERENCES

1. S. Raju, P. Mohammed Akhtar, *"Time Series Analysis on Rainfall and Ground Water Levels Data – A Case Study"*, International Journal of Scientific Research in Mathematical and Statistical Sciences (IJSRMSS) Vol.6, Issue.1, pp.76-85, February(2019).
2. S. Raju, P. Mohammed Akhtar, *"Fitting of modified exponential model between rainfall and ground water levels: A case study"*, International Journal of Statistics and Applied Mathematics 2019; 4(4), pp. 01-06.
3. S. Raju, P. Mohammed Akhtar, *"Fitting Of Gompertz Model Between Rainfall And Ground Water Levels – A Case Study"*, International Journal of Mathematics Trends and Technology (IJMTT) – Volume 65, Issue 7 – July (2019), pp. 85-93.
4. S. Raju, P. Mohammed Akhtar, *"Fitting of Logistic Model between Rainfall and Ground Water Levels – A Case Study"*, Compliance Engineering Journal, Volume 10, Issue 9, pp. 114-122, September-2019.
5. S. Raju, P. Mohammed Akhtar, *"Fitting of Gompertz model between Rainfall and Ground water levels by using Partial sums method – A case study"*, Stochastic Modeling and Applications, Vol. 25 No.1 (January-June, 2021), MuK Publications and Distributions (INDIA). ISSN: 0972-3641.
6. S. Raju, *"Fitting of Modified Exponential Model between Rainfall and Ground Water Levels by Using Partial Sums Method – A Case Study"*, Pensee Journal, Volume 51, Issue 03, pp. 647-655, March-2021.