551·577(54)

Correlation between mean annual rainfall and its standard deviation

V. N. ANTARKAR

Meteorological Office, Poona

(*Received 14 December 1961*)

ABSTRACT. The hypothesis that linear relationship exists between mean rainfall and its standard deviation is tested in case of Rajasthan by obtaining correlation coefficient between them and developing linear regression. It is seen that high correlation exists and that the observed differences between the estimated and actual standard deviation are not, on the basis of Standard Error of estimate, significant. Relative error in the estimate is also generally small.

1. It is an observed fact that Coefficients of Variability CV generally varies inversely with rainfall (particularly in case of mean annual rainfall). Hence, we may have as a first approximation,

$$CV = a + b/x,$$

where x is the mean rainfall of a station and a and b are constants. But by definition $CV = s/x$, where s is the standard deviation.

$$\therefore\ s/x = a + b/x \qquad \text{or} \quad s = ax + b$$

That is a priori, linear relationship between x and s is suggested.

2. This hypothesis was tested in case of Rajasthan for which values of s for over 190 stations are easily available (Rao 1958). For purposes of this study, Rajasthan has been divided into two regions—(1) the region R_1, containing stations with mean annual rainfall of less than 22″ and (2) the region R_2, with stations having mean annual rainfall of 22″ and over. These two regions roughly correspond to the meteorological subdivisions of west Rajasthan and east Rajasthan. The number of stations, considered here, is 87 in R_1 and 106 in R_2.

3. The correlation coefficient r and the regression equation for each region were obtained considering values of x and s for each station. Standard Error, E_s, of the estimate s_e was also calculated from —

$$E_s = \sigma_s \, (1 - r^2)^{\frac{1}{2}} ,$$

where σ_s is the standard deviation of the s values. These equations and values are as follows —

For Region R_1

$$s_e = 0 \cdot 310x + 2 \cdot 51$$

$$r = 0 \cdot 85, \ F_s = 1 \cdot 00''$$

For Region R_2

$$s_e = 0 \cdot 211\, x + 4 \cdot 77$$

$$r = 0 \cdot 61, \quad E_s = 1 \cdot 60''$$

It is interesting to note that both the equations are almost satisfied when $x = 22 \cdot 83$ which closely corresponds to the isohyet of 22″ dividing the two regions.

3.1. Estimated standard deviation s_e was calculated for each station from the regression equations. The difference $(s_e - s_a)$ between the estimated and actual standard deviations was obtained. Frequency distribution of this error for each region is given in Table 1 in terms of the standard error. On the basis of the normal distribution of this error, we could expect it to lie between

$\pm 0 \cdot 67\, E_s$ in case of 50% of the stations,

$\pm 1 \cdot 0\, E_s$ in case of 68·2% of the stations,

$\pm 2 \cdot 0\, E_s$ in case of 95·5% of the stations, and

$\pm 2 \cdot 58\, E_s$ in case of 99% of the stations.

TABLE 1

Actual and percentage number of stations with error within E_s times

Error $(s_e - s_a)$	$R_1(E_s = 1 \cdot 00'')$		$R_2(E_s = 1 \cdot 60'')$	
	Actual	Per cent	Actual	Per cent
$<-2 \cdot 58$	38	43·7	45	42·5
−2·58	37	42·5	43	40·6
−2·50	37	42·5	42	39·6
−2·25	37	42·5	42	39·6
−2·00	37	42·5	42	39·6
−1·75	36	41·4	41	38·7
−1·50	33	37·9	35	33·0
−1·25	31	35·6	32	30·2
−1·00	26	30·0	27	25·5
−0·75	20	23·0	22	20·8
−0·67	19	21·8	22	20·8
−0·50	17	19·5	17	16·0
−0·25	7	8·0	9	8·5
0	0	0	0	0
0·25	10	11·5	17	16·0
0·50	21	24·1	29	27·4
0·67	29	33·3	39	36·7
0·75	29	33·3	43	40·6
1·00	35	40·2	49	46·2
1·25	40	46·0	57	53·8
1·50	45	46·0	58	54 7
1·75	48	55·2	60	56·6
2·00	49	56·3	60	56·6
2·25	49	56·3	61	57·5
2·50	49	56·3	61	57·5
2·58	49	56·3	61	57·5

The number of stations falling within these limits is also indicated in the table. The following extract from Table 1 indicates how closely these expectations have been realised—

Actual and percentage number of stations with error within

	$0 \cdot 67\, E_s$	$1 \cdot 0\, E_s$	$2 \cdot 0\, E_s$	$2 \cdot 58\, E_s$	
Actual	48	61	86	86	R_1
%	55·2	70·1	98·9	98·9	
Actual	61	76	102	104	R_2
%	57·5	71·7	96 2	98·1	

The relative percentage error $100\,(s_e - s_a)/s_a$ was also calculated for each station. Table 2 gives the distribution of stations according to the relative error.

4. Next, the stations in each region were grouped according to convenient and small rainfall intervals — range of each interval being $0 \cdot 75''$ in case of R_1 and $1 \cdot 00''$ in case of R_2. The mean rainfall for each interval was obtained to give the area-mean rainfall for the area covered by each interval. The best estimate of the standard deviation for this mean spatial rainfall was obtained as

$$\hat{s} = \left(\sum_{1}^{k} n_i \; s_i^2 \Big/ (N - k) \right)^{\frac{1}{2}}$$

where n_i is the number of years of data of a station i used for getting S.D., s_i of that station, $N = \sum_{1}^{k} n_i$ and k is the number of stations. Tables 3 and 4 give these values. Maldistribution of raingauge stations is also incidentally brought out. The correlation coefficients between x and $\hat{s}$, the standard errors and the regression equations were obtained and are given below—

Region R_1

$$\hat{s}_e = 0 \cdot 343x + 2 \cdot 06$$

$$r = \cdot 94, \; E_{\hat{s}} = 0 \cdot 68$$

TABLE 2

Actual and percentage number of stations with relative % error between

	0—10	11—20	21—30	31—40	41—50	51—60	61—70	71—80	81—90	
Actual	48	23	9	4	1	1	0	0	1	R_1
%	55·2	26·5	10·3	4·6	1·1	1·1	0	0	1·1	
Actual	64	31	9	1	1					R_2
%	60·4	29·2	8·5	1	1					

TABLE 3

Region R_1

Rainfall range (cents)	No. of stations	Mean rainfall x (cents)	$\hat{s}$ (cents)
326— 400	1	349	195
401— 475	4	450	207
476— 550	3	520	400
551— 625	3	596	417
626— 700	2	676	555
701— 775	6	737	517
776— 850	3	798	525
851— 925	3	907	482
926—1000	2	935	527
1001—1075	5	1053	585
1076—1150	7	1120	685
1151—1225	3	1201	632
1226—1300	1	1233	560
1301—1375	2	1362	703
1376—1450	5	1422	708
1451—1525	4	1498	776
1526—1600	4	1548	819
1601—1675	5	1634	757
1676—1750	2	1734	831
1751—1825	3	1780	865
1826—1900	6	1863	898
1901—1975	3	1966	765
1976—2050	4	2017	801
2051—2150	3	2076	866
2151—2200	3	2180	935

TABLE 4

Region R_2

Rainfall range (cents)	No. of stations	Mean rainfall x (cents)	$\hat{s}$ (cents)
2201—2300	13	2251	1025
2301—2400	7	2339	1001
2401—2500	7	2442	979
2501—2600	11	2538	1042
2601—2700	9	2644	971
2701—2800	6	2769	966
2801—2900	4	2857	1058
2901—3000	5	2948	1094
3001—3100	5	3079	1271
3101—3200	4	3154	1360
3201—3300	3	3258	1300
3301—3400	5	3356	1306
3401—3500	6	3469	1375
3501—3600	1	3553	1206
3601—3700	3	3644	1214
3701—3800	6	3739	1254
3801—3900	4	3870	1305
3901—4000	3	3955	1240
4001—4100	0	—	—
4101—4200	1	4111	1335
4201—4300	1	4237	1618
4301—4400	1	4305	1312
4401—4500	1	4443	1502

TABLE 5

Actual and percentage number of intervals with % error between

	0—10	11—20	21—30	31—40	41—50	51—60	61—70	71—80	
Actual	17	5	1	0	0	0	1	1	R_1
%	68	20	4	0	0	0	4	4	
Actual	19	3							R_2
%	86·4	13·6							

Region R_2

$$\hat{s}_e = 0 \cdot 224x + 4 \cdot 72$$

$$r = 0 \cdot 84, \ \hat{E}_s = 0 \cdot 95$$

Here again, both the equations are almost satisfied by x=22·35, which closely corresponds to the isohyet of 22″, dividing the two regions.

4.1. In this case also, estimated standard deviation $\hat{s}_e$ was calculated for each interval (representing a uniform rainfall region because of narrow isohyetal boundaries). The difference $(\hat{s}_e - \hat{s}_a)$ between the estimated and actual standard deviations was obtained.

Table 6 gives the distribution and shows how closely the expectations indicated in para 3.1 have been realised.

The relative percentage error $100 \ (\hat{s}_e - \hat{s}_a) / \hat{s}_a$ was also calculated. Table 5 gives distribution of intervals according to the relative error. It is seen that although the estimated standard deviation $\hat{s}$ may be out by more than one $\hat{E}_s$ in case of about 30% of the intervals the relative error is quite small being generally less than 20%.

TABLE 6

Actual and percentage number of intervals with error within

	$0 \cdot 67\hat{E}_s$	$1 \cdot 0\hat{E}_s$	$2 \cdot 0\hat{E}_s$	$2 \cdot 58\hat{E}_s$	
Actual	12	17	24	25	R_1
%	48·0	68·0	96·0	100·0	
Actual	12	15	21	22	R_2
%	54·5	68·2	95·4	100·0	

5. Wide variations of standard deviation from observed values could be attributable to (1) peculiarities of orography and / or defective exposure conditions, apart from meagreness of data in some cases. Nevertheless, the study has shown that a fairly close relationship exists between rainfall and its standard deviation for individual stations as well as for regions of homogeneous rainfall (bounded by isohyets at narrow intervals); and the observed differences are not statistically significant. Hence, isohyetal map of any region can be used as a map giving distribution of standard deviation.

REFERENCE

Rao, K. N. 1958 *Indian J. Met. Geophys.*, 9, 2, p. 99.