551.582.2
311.14 : 311.15

CLIMATOLOGICAL NORMALS

By J.V. Maher

Central Office, Bureau of Meteorology, Melbourne

(Manuscript received July 1966)

ABSTRACT

A brief examination is made of the current standard practices regarding normals, period averages and climatological normals. The standard error of the mean and the effect of sample size in the case of normal distributions and also for the type of distribution met in practice in Australia are considered for monthly and annual rainfall totals. The variations in ratios of rainfall normals were also examined and the conclusion is reached that the use of a comparatively short period (30 years) for quintiles and normals is often not satisfactory and that where there are no secular changes in climate, the longest possible period should be used.

1. DEFINITIONS

The Arithmetic Mean or Average of a number of observations $x_1, x_2, \ldots\ldots x_n$ is defined as,

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i , \quad \text{where } i = 1, 2, \ldots\ldots n$$

Period Averages: Averages of climatological data computed for any period of at least ten years starting on 1 January of a year ending with the digit 1.

Normals: Period averages computed for a uniform and relatively long period comprising at least three consecutive 10-year periods.

Climatological Standard Normals: Averages of climatological data computed for the following consecutive periods of 30 years: 1 January 1901 to 31 December 1930, 1 January 1931 to 31 December 1960, etc. (When data are not continuous, adjusted normals may be computed).

In a note to WMO Technical Regulation 8.4.2.5., it is pointed out that for stations from which records are not available for the computation of period averages, normals or climatological standard normals, averages for shorter periods (e.g. five years) may be useful, for example, in tropical countries, for ocean weather stations and for upper-air stations.

2. STANDARD ERROR OF THE MEAN AND THE EFFECT OF SAMPLE SIZE

It can be shown that the standard deviation (or standard error) of the distribution of the means of a series of random samples of size n from a parent normal distribution is $\sigma / n^{\frac{1}{2}}$, where σ is the standard deviation of the parent population.

Where the value of σ is not known, as is often the case with meteorological populations, it is necessary to use s, the best estimate of σ.

The best estimate of the standard error of the distribution of $\bar{x}$ then becomes $s/n^{\frac{1}{2}}$.

Rofe, B. 1963 "Australian sounding rocket experiments". Technical Note SAD 127, Department of Supply, Salisbury, Australia.

Sparrow, J.G. and Unthank, E.L. 1964 "Biennial stratospheric oscillations in the Southern Hemisphere". J. Atmos. Sci., 21, p. 592.

Tucker, G.B. 1964 "Zonal winds over the Equator". Quart. J.R. Met. Soc., 90, p. 405.

1965a "The divergence of horizontal eddy flux of momentum in the lower equatorial stratosphere". Quart. J.R. Met. Soc., 91, p. 356;

1965b "Comments on 'An evaluation of certain forms of momentum transfer in the 26-month oscillation'". Correspondence, Mon. Weather Rev., 93, p. 704.

Wallace, J.M. and Newell, R.E. 1965 "Eddy fluxes and the biennial oscillation in the stratosphere". Report No. 14, M.I.T. Department of Meteorology.

From this it is possible to derive confidence limits for the true value of the population mean μ, based on a sample mean $\bar{x}$ of n values of the population.

The 95% confidence limits of μ, for example, are $\bar{x} \pm (t_{.05,n} \cdot s/n^{\frac{1}{2}})$, where $t_{.05,n}$ is the 5 per cent level of Students t for n-1 degrees of freedom. The variation of t with (n-1) at the 5 per cent level is shown below.

n-1	20	24	30	40	60	120
t	2.09	2.06	2.04	2.02	2.00	1.98

The expression for the confidence limits is thus approximately,

$$\bar{x} \pm 2\ s/n^{\frac{1}{2}}$$

and the effect, for example, of increasing n from 25 to 100 would be to decrease the confidence limits by half.

It should be stressed that the above theory applies to random samples from a normal population each element of which is independent of the other.

In Australia, annual rainfalls are generally normally distributed but in many places monthly rainfall distributions are heavily skewed by the occurrence of a preponderance of very small or zero rainfalls combined with a relatively few heavy falls.

Table 1 shows the effect of sample size on the 95% confidence limits of averages (normals) or annual rainfall at four Australian stations. Melbourne and Caulfield are stations in a reliable rainfall region, whilst Mardie and Onslow have been selected to represent a region of high variability.

Table 1. Annual rainfall statistics for four Australian stations

Station	Period	$\bar{x}$ (in.)	s (in.)	n (yr)	95% Conf. Limits (in.)	
					Lower	Upper
MELBOURNE	1931-1960	27.21	4.95	30	25.37	29.05
37° 49'S, 144° 58'E	1856-1964	25.88	4.95	109	24.93	26.83
CAULFIELD	1931-1960	30.67	5.94	30	28.46	32.88
37° 54'S, 145° 00'E	1888-1964	28.66	5.94	77	27.31	30.01
MARDIE	1931-1960	11.19	6.31	30	8.84	13.54
21° 12'S, 115° 57'E	1886-1964	10.40	6.31	77+	8.95	11.85
ONSLOW	1931-1960	10.79	7.77	30	7.90	13.68
21° 40'S, 115° 07'E	1886-1964	10.40	7.77	79	8.65	12.15

+ 1887 and 1890 missing

3. ACTUAL VARIATIONS IN RAINFALL NORMALS IN AUSTRALIA

(a) Quintiles

Where rainfall is approximately normally distributed it is convenient to express the degree of departure from average (or normal) conditions by reference to quintiles as shown below:-

Rainfall in the 1st Quintile Range - well below average
" " " 2nd " " - below average
" " " 3rd " " - average
" " " 4th " " - above average
" " " 5th " " - well above average

However, in many places in Australia there are months in which it is rare for any rainfall to occur. Here zero rainfalls occur in the 5th quintile range and the anomaly arises that under the system above, these months would be described as 'well above average'.

(b) Effect of Skewness and Runs

The skewed distributions also strongly affect short period averages and 30-year normals because there is a sharp change in the average when one of the occasional heavy rainfalls occurs in the period used. If the average is calculated each year, a saw tooth pattern is achieved, with sharp rises after a rare wet month, followed by gradual decrease of the average with successive years of zero or very low rainfall in the particular month.

Longer period records are less affected, but an average calculated from a few months with heavy falls and numerous completely dry months has little reality.

Longer period normals also vary considerably due to long runs of below or above average rain or short runs of large departures from normal. Some statistics of runs for Melbourne and Mardi are given in Table 2.

Table 2. Frequency of occurrence of runs of annual rainfalls above and below average

Station	Period of Record	No. of Years Run Above Average							No. of Years Run Below Average						
		8	7	6	5	4	3	2	8	7	6	5	4	3	2
		Frequency							Frequency						
Melbourne	1856-1964	1	0	0	0	0	4	10	0	0	0	1	4	4	3
Mardie	1886-1964+	0	0	0	1	2	2	5	1	0	0	1	0	2	4

\+ 1887 and 1890 missing

(c) Variation of 30-year Rainfall Normals

For the four stations Melbourne, Caulfield, Mardie and Onslow, all of the possible 30-year annual rainfall normals were calculated (overlapping normals) and the frequency of occurrence of the values so obtained are given in Tables 3 to 5. From the Melbourne record, 80 such values were obtained, for Caulfield 48, for Mardie 40 and for Onslow 50. In addition 25 sets of 30 randomly selected annual rainfalls for Melbourne and Mardie and the 30-year annual normals calculated for each set. The distribution of these values is also given in Tables 3 and 5.

Table 3. Frequency of occurrence of values of 30-Year Annual Rainfall Normals at Melbourne

Range (Inches)	24.25 - 24.74	24.75 - 25.24	25.25 - 25.74	25.75 - 26.24	26.25 - 26.74	26.75 - 27.24	27.25 - 27.74	Total
Melbourne (Running Annual)	4	17	28	21	1	9	0	80
Melbourne (Random Annual)	3	5	2	6	4	3	2	25

Table 4. Frequency of occurrence of values of 30-Year Annual Rainfall at Caulfield

Range (Inches)	26.75 - 27.24	27.25 - 27.74	27.75 - 28.24	28.25 - 28.74	28.75 - 29.24	29.25 - 29.74	29.75 - 30.24	30.25 - 30.74	30.75 - 31.24	Total
Caulfield (Running Annual)	21	13	2	2	1	0	3	5	1	48

Table 5. Frequency of occurrence of values of 30-Year Annual Rainfall at Mardie and Onslow

Range (Inches)	8.00 - 8.49	8.50 - 8.99	9.00 - 9.49	9.50 - 9.99	10.00 - 10.49	10.50 - 10.99	11.00 - 11.49	11.50 - 11.99	12.00 - 12.49	Total
Mardie (Running Annual)	0	2	9	6	6	5	8	3	1	40
Mardie (Random Annual)	0	1	5	7	3	4	3	2	0	25
Onslow (Running Annual)	3	5	6	9	10	9	4	3	1	50

To illustrate the very wide scatter for individual months, the frequency of occurrence in various ranges is shown also for Mardie (Jan., Nov.) and for Melbourne (Jan., July) in Tables 6 and 7 respectively. All possible 30-year monthly rainfall normals were used.

Table 6. Frequency of occurrence of values of 30-Year Rainfall Normals for January and November at Mardie.

January

Range (Inches)	.80 - .99	1.00 - 1.19	1.20 - 1.39	1.40 - 1.59	1.60 - 1.79	1.80 - 1.99	2.00 - 2.19	2.20 - 2.39	Total
Mardie (Running) January	0	23	8	5	2	1	1	0	40
Mardie (Random) January	3	4	7	3	6	1	0	1	25

November

Range (Inches)	.01 - .02	.03 - .04	.05 - .06	.07 - .08	.09 - .10	.11 - .12	.13 - .14	.15 - .16	Total
Mardie (Running) November	6	8	4	0	6	3	9	4	40
Mardie (Random) November	1	4	0	11	4	4	1	0	25

Table 7. Frequency of occurrence of values of 30-Year Rainfall Normals for January and July at Melbourne

January

Range (Inches)	13.0 - 13.9	14.0 - 14.9	15.0 - 15.9	16.0 - 16.9	17.0 - 17.9	18.0 - 18.9	19.0 - 19.9	20.0 - 20.9	21.0 - 21.9	22.0 - 22.9	23.0 - 23.9	Total
Melbourne (Running) January	(1) 0	(2) 0	(3) 0	(4) 2	(5) 12	(6) 21	(7) 24	(8) 14	(9) 7	(10) 0	(11) 0	(12) 80
Melbourne (Random) January	1	0	1	2	5	6	4	3	2	0	1	25

Table 7. (Cont'd)

July

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Melbourne (Running) July	0	0	0	0	4	7	5	5	4	0	0	25
Melbourne (Random) July	0	0	0	1	17	16	32	8	6	0	0	80

The frequencies in Tables 3 to 7 show that, particularly for monthly 30-year rainfall normals, there is a wide range of possible values. For example, the 30-year normal rainfall for Mardie for November varies from 0.01 inches to 0.15 inches according to the period used. The frequencies in many cases are not symmetrically distributed because for random selections the number taken is too few and for the running 30-year normals the values are strongly affected by the grouping of the light and heavy falls. Where there are only rare heavy falls the normals including them are large and those excluding them are small and there are few intermediate values.

4. RAINFALL RATIOS USING 30 – YEAR AND LONG PERIOD NORMALS

The principal and obvious advantage of standard period normals is for comparisons of conditions at adjacent stations. To investigate the basis of this, the ratios of annual rainfall normals between Melbourne and Caulfield (7 miles apart) and Mardie and Onslow (30 miles apart) have been examined and are tabulated below. Using running 30-year normals from the period 1896-1925 to the period 1935-1964, the ratio Mardie/Onslow for annual rainfall varied from .924 for the 30 years 1900-1929 to 1.055 for the 1928-1957 period. Frequencies of occurrence of the various ratios are given in Table 8.

Table 8. Frequency of Occurrence of Ratios of Pairs of Running 30-Year Annual Rainfall Normals at Mardie and Onslow

Ratio	.920 - .939	.940 - .959	.960 - .979	.980 - .999	1.000 - 1.019	1.020 - 1.039	1.040 - 1.059	1.060 - 1.079	Total
Freq.	3	5	11	9	4	3	5	0	40

For Melbourne and Caulfield the records were available for both stations from 1888-1964 and 48 pairs of 30-year normals were calculated using data for overlapping 30-year periods from 1888-1917 to 1935-1964. The ratio Melbourne/Caulfield for annual rainfall varied from .868 for 1935-1964 to .958 for the 1912-1941 period. Frequencies of occurrence of the various ratios are given in Table 9.

Table 9. Frequency of Occurrence of Ratios of Pairs of Running 30-Year Annual Rainfall Normals at Melbourne and Caulfield

Ratio	.860 - .879	.880 - .899	.900 - .919	.920 - .939	.940 - .959	Total
Freq.	2	7	8	11	20	48

One characteristic of running means is that although all possible sets have been used they are not independent and values derived do not form a true statistical distribution. Also in any set the middle of the period has a weighted influence. For example, in the 48 sets of running 30-year normals from 1888-1917 down to 1935 to 1964, the year 1888 is used once, 1889 twice and so on up to the years 1917 to 1935 which are each used 30 times. From 1936 which is used 29 times, the useage decreases successively to once for 1964. The frequencies in Table 9 for Melbourne/Caulfield were strongly influenced by the ratios arising out of the sets of pairs from 1900-1929 to 1919-1948 when the ratios were all in the range from .940 to .959.

If all years of record are used and the periods are reasonably long the normals are more stable and this is shown (see Table 10) by averages for all years to 1920, 1930, 1940, 1950 and 1960, for Melbourne, Caulfield, Mardie and Onslow.

Table 10. Statistics based on all years of record ending in various years

Period All Years to	Melbourne	Caulfield	Ratio Melbourne/Caulfield	Mardie	Onslow	Ratio Mardie/Onslow
1920	25.61	27.58	.929	8.17	8.48	.963
1930	25.51	27.02	.944	8.54	9.05	.944
1940	25.57	27.28	.937	8.71	9.16	.951
1950	25.62	27.50	.932	9.40	9.68	.971
1960	25.99	28.53	.911	9.63	9.74	.989

The variability of the ratios of the rainfall averages is not surprising when the theory of the distribution of the ratios of two random variables is considered, noting as before that the overlapping 30-year normals are not true random variables.

Consider the distribution of the ratio z of two random variables x and y, where $z = x/y$ and $y \neq o$ (see, e.g., Kendall and Stuart, 1958).

Now $dz = \frac{\partial z}{\partial x} dx + \frac{\partial z}{\partial y} dy$ at any point in the x, y, z space and since $\log_e z = \log_e x - \log_e y$

$$\frac{\partial z}{\partial x} = \frac{z}{x} = \frac{1}{y}, \quad \frac{\partial z}{\partial y} = \frac{-z}{y} = \frac{-x}{y^2}$$

Therefore $dz = \frac{1}{y} dx - \frac{x}{y^2} dy$ and squaring we have,

$$dz^2 = \frac{(dx)^2}{y^2} + \frac{x^2}{y^4}(dy)^2 - \frac{2x}{y^3} dx\ dy .$$

If the departures are considered to be measured from their means, we have, taking sums and means,

$$V(z) = \frac{V(x)}{\mu_y^2} + \frac{\mu_x^2}{\mu_y^4} . V(y) - 2\frac{\mu_x}{\mu_y^3} \text{Cov}(x, y)$$

Where $V(z)$ is the variance of z,

μ_x is the population mean of x,

μ_y is the population mean of y,

$\text{Cov}(x, y)$ is the population covariance of x and y .

Note that because of the substituion of mean values for the partial derivatives, the above expression is not exact but is true when the population (sample) is large.

Since $\text{cov}(x, y) = (V(x).V(y))^{\frac{1}{2}} \rho(x, y)$

where ρ is the population value of the correlation coefficient, the expression for $V(z)$ may be written

$$V(z) = \frac{1}{\mu_y^2} \left[V(x) + \frac{\mu_x^2}{\mu_y^2} V(y) - 2 \frac{\mu_x}{\mu_y} . V(x)^{\frac{1}{2}} V(y)^{\frac{1}{2}} \rho(x, y) \right] .$$

In the case of a sample from a population we must write the equation thus:

$$s^2(z) = \frac{1}{\bar{y}^2} \left[s^2(x) + \frac{\bar{x}^2}{\bar{y}^2} s^2(y) - 2 \frac{\bar{x}}{\bar{y}} . s(x) . s(y)\ r(x, y) \right]$$

where $s^2(z)$ represents the sample value of $V(z)$ etc., and $r(x, y)$ is the sample value of the correlation coefficient. It is obvious from the above equation that the correlation coefficient plays a major role in determining the magnitude of the variance of the ratio z. For example, if (s $(x) \simeq s(y)$ and $\bar{x} \simeq \bar{y}$, the dependence of the value of the term inside the bracket on the value of r is more evident. If $r \simeq 1$, this term will be close to zero. If, on the other hand $r = -1.0$, the term inside the square bracket is approximately equal to $4\ s^2(x) \simeq 4s^2(y) \simeq 4\ s(x)\ s(y)$. Thus for ratios between annual totals or means of groups of rainfall totals to be reasonably constant and thus meaningful, particularly when used singly (as in standard 30-year rainfall normals), the correlation coefficient should be positive and as large as possible. Recent studies which have been made in connection with drought have shown that, in general, correlation between two rainfall series falls away rapidly with increasing separation of the two rainfall stations.

5. CONCLUSIONS

It is clear from the above that ratios and normals from the longer period records are more stable and that the use of comparatively short period records (30 years) for rainfall normals is often not satisfactory in Australia, and this applies particularly to normals derived for single months. The stations used are representative of large areas of Australia and it is concluded that where there are no secular changes in climate, the longest possible period should be used.

Published standard period normals have value only for approximate comparisons between adjoining stations, the short standard period used (30 years) ensuring the possibility of the publication of a large number of stations.

ACKNOWLEDGMENT

I wish to thank Dr. H.C.S. Thom for his comments on this paper when it was originally submitted as document No. 52 to the W.M.O. Commission for Climatology, 4th Session in Stockholm, 1965. Following the discussions with Dr. Thom the section on the variance of the ratio of two random variables was added.

REFERENCES

Kendall, M.G. and Stuart, A. 1958 The Advanced Theory of Statistics, Vol. 1, p. 231. Charles Griffin and Company Ltd., London.