551.501.45:311.16

THE SIGNIFICANCE OF CORRELATIONS IN METEOROLOGICAL SERIES

by G. O'Mahony

Central Office, Bureau of Meteorology, Melbourne

(Manuscript received April 1960)

Abstract: The definition of the correlation coefficient and how the significance of a value derived from a sample is dependent on the number of pairs used in the computation is set out. The complications introduced by the presence of serial correlation within the parent series is brought out. A survey is given of the recent statistical literature which is relevant, and in which is contained a good coverage of the methods for testing significance, both of the correlation between serially correlated series, and also of the serial correlation coefficients themselves. Finally, suggestions are made as to the particular method of testing to be used in practice, the ultimate selection depending on interim results and the degree of precision required.

1. INTRODUCTION

The correlation coefficient between two series

x_1, x_2 --------- x_n and

y_1, y_2 --------- y_n is computed by

$$r_{xy} = \sum_1^n (x - \bar{x})(y - \bar{y}) \left[\sum_1^n (x - \bar{x})^2 \sum_1^n (y - \bar{y})^2\right]^{-\frac{1}{2}}$$

where $\bar{x}$ and $\bar{y}$ are the mean values of the two samples and n is the number of pairs used in the calculation.

The value of r found in this way is an indication of how the two series behave with respect to each other, and to this stage, the result may be classed as purely qualitative. In practice this is usually not sufficient, and what is usually needed is a test of significance of the value obtained. The experimenter seeks an answer to the question - "Does this value of $|r|$ really represent a correlation between the two series, or could it (or a larger value) have arisen by chance, i.e. is the result likely to have been caused by sampling fluctuation?"

The method of answering this question is to adopt the null hypothesis that there is no correlation between the series, i.e. the population value (ρ) is equal to zero. Then, if n is large and $|r|$ not almost equal to unity, the distribution of r can be assumed to be normal with a standard error of $(n-1)^{-\frac{1}{2}}$. If the observed value of $|r|$ is $\geq$ 2 S.E. it is significant at the 5% level, and so the null hypothesis would be rejected. With n small or with $|r|$ nearly unity the distribution of r is not approximately normal and it is necessary to apply another test which involves Student's t.

Weatherburn (1947) shows that for random samples of n pairs of values from an uncorrelated bivariate normal population, r^2 is a Beta variate of the first kind with parameters $\frac{1}{2}$ and $\frac{1}{2}(n-2)$. The Beta variate of the first kind is defined as a continuous variate, say x, which is distributed with probability density

$$\phi(x) = \left[x^{l-1}(1-x)^{m-1}\right]\left[\beta(l,\ m)\right]^{-1}$$

where $\beta(l, m)$ is the β function and is given by

$$\beta(l,\ m) = \int_0^1 x^{l-1}\ (1-x)^{m-1}\ dx$$

Since r^2 is a $\beta_1\left(\frac{1}{2}, \frac{1}{2}(n-2)\right)$ variate then

$r^2\ (1-r^2)^{-1}$ is a $\beta_2\left(\frac{1}{2}, \frac{1}{2}(n-2)\right)$ variate where $\beta_2\left(\frac{1}{2}, \frac{1}{2}(n-2)\right)$ is a Beta variate of the second kind which defines the distribution of a continuous variate, say y, as

$$\phi(y) = y^{-\frac{1}{2}}\ (1+y)^{-\frac{1}{2}(n-1)}\ \left[\beta\left(\tfrac{1}{2},\ \tfrac{1}{2}(n-2)\right)\right]^{-1}$$

where the range is now from 0 to ∞.

Writing $t = r\ (n-2)^{\frac{1}{2}}\ (1-r^2)^{-\frac{1}{2}}$

makes

$t^2\ (n-2)^{-\frac{1}{2}}$ a $\beta_2\left(\frac{1}{2}, \frac{1}{2}(n-2)\right)$ variate and so $r\ (n-2)^{\frac{1}{2}}\ (1-r^2)^{-\frac{1}{2}}$ is distributed as Student's t with (n-2) degrees of freedom.

To test whether an observed value of $|r|$ is likely to have occurred by chance, the t value is calculated by substitution of the relevant n and r values, and t - tables are consulted to ascertain the probability of occurrence of such a limiting value.

Because the significance value of $|r|$ is a function of n the number of pairs used in its calculation, there is obviously little value in comparing two sample values of r without reference to the sample size. Thus, R.A. Fisher (1948) has computed tables of minimum values of r that are significant at the 5% level, and some extracts from his tabulations are:- for n = 6 the appropriate value of r is 0.811; for n = 12, the relevant r value is 0.576; for n = 100, the r value is approximately 0.192.

2. CORRELATIONS OF METEOROLOGICAL SERIES

The tests described above are valid if, and only if, the basic assumption is fulfilled that each pair of values x_i, y_i used in the calculation is completely independent of every other pair. It is here that most confusion has arisen in the testing of correlations between meteorological series, where in most cases the assumption of independence is not justified.

If, for example, the x_1, x_2 ----- x_n represent daily values of 700 mb temperatures at one station, and the y_1, y_2 ----- y_n are 700 mb temperatures on respective days at another station, then quite obviously a value x_t is generally not independent of the values x_{t-1} or x_{t+1}, and similarly the value y_t is not independent of the y values immediately before or after it. Where the series are not independent in this way, they are said to be serially correlated, and the tests described in the previous section are invalid.

An approximate illustration of this is furnished by considering the basis of the test using the standard error of the sample value of r. There, the standard error was taken to be $(n-1)^{-\frac{1}{2}}$ where the value of n refers to the number of independent pairs. Then, in a series of 900 pairs, assume that the number of independent pairs is only 100. The apparent value of the standard error is approximately 0.03, i.e. assuming all 900 pairs are independent, but the true value of the SE. is approximately 0.10. An observed value of $|r|$ of say 0.09 computed from 900 independent pairs would be equal to 3 x SE. and so would be highly significant, but the same value arising from the second case would be less than one standard error and so would be well within the range of chance occurrences.

The recent statistical literature has given a lot of space to this difficulty in assessing significance of correlations between time series, but the meteorological journals have given it surprisingly little publicity, considering that the bulk of meteorological data come under the category of belonging to serially correlated series.

3. TESTING OF SIGNIFICANCE OF CORRELATION BETWEEN TWO SERIALLY CORRELATED SERIES

The testing of the significance of correlation between two non-independent series, like the testing of the serial correlations themselves which is considered later, is a very complex problem. The remainder of this paper will be devoted to providing an outline of the modern methods of attacking this problem, and so is really a survey of the relevant literature.

Bartlett (1935) in illustrating the effects of serial correlation within two parent series which are not correlated together, computes the variance of the sample correlation between the two series to be approximately

$$n^{-1} (1 + \rho_1 \rho_2) (1 - \rho_1 \rho_2)^{-1}$$

where n is the number of pairs used in the calculation,

ρ_1 is the serial correlation between adjacent terms in one series, and

ρ_2 is the serial correlation between adjacent terms in the other series.

Then is ρ_1 and ρ_2 have the same sign, there is less accuracy in the estimation of the correlation than there would be if the pairs were independent. In the example let $\rho_1 = \rho_2 = 0.6$, then the accuracy realised in estimating a correlation which is approximately zero would be equivalent to that derived from less than half the same number of independent pairs. If pairs were independent the variance of the correlation coefficient would be approximately n^{-1}, but in the serially correlated case quoted, the variance would be approximately $(n^{-1}) (1.36) (0.64)^{-1}$ and this is greater than $2 n^{-1}$.

Moran (1947) computes the standard error of the coefficient of correlation between two such series to be given approximately by $\left\{ n^{-1} \left(1 + 2 \sum_1^{\infty} \rho_s \bar{\rho}_s\right) \right\}^{\frac{1}{2}}$, where n is the length of the series, and ρ_s and $\bar{\rho}_s$ are the serial correlation coefficients of the two series. This will be seen to be approximately equal to the value obtained by Bartlett (the difference for s = 1 is $2 (\rho_1 \bar{\rho}_1)^2 n^{-1}$). The problem is not solved at this stage, since Bartlett's ρ_1 and ρ_2 and Moran's ρ_s and $\bar{\rho}_s$ are population values and their estimation from samples is extremely complicated. In his paper Moran shows that testing for independence between two serially correlated series really depends on the type of probability processes which generate the series.

He takes two stationary series $X_{(t)}$ and $Y_{(t)}$ with zero means and finite moments up to and including the fourth, proves that under certain assumptions the covariance $S = \sum_{1}^{n} X_{(t)} Y_{(t)}$ tends to normality as n increases, and shows that under the same conditions

$$E(S) = 0$$

$$E(S^2) = n C_o d_o + 2 \sum_{s=1}^{n-1} (n-s)\, C_s\, d_s$$

where
$$C_o = E\left[X(t)^2\right], \qquad d_o = E\left[Y(t)^2\right]$$
$$C_s = E\left[X(t).X(t-s)\right], \quad d_s = E\left[Y(t).Y(t-s)\right]$$

In practice the application of this theory is further complicated by the fact that sample covariances cannot be used in the calculation of the expected value of S^2, but the C_o and d_o must be computed in terms of the coefficients of the stochastic difference equation. Thus if the X series is of the form

$$X_{(t+2)} + a\, X_{(t+1)} + b\, X_{(t)} = \eta_{(t+2)}$$

then $C_o = \sigma_1^2 \dfrac{1 + b}{(1-b)\left\{(1+b)^2 - a^2\right\}}$ where σ_1^2 is variance of η.

A similar relationship exists for the d_o and the Y series provided of course this series is of the same form.

In a later paper Moran (1949) refers to the same problem and its attendant difficulties. He states that as generally the presence of serial correlation causes the standard error of the correlation coefficients between series to be larger than it would be if successive pairs were independent, then a value which is found to be not significant, assuming independence, is unlikely to be judged significant if the more exact formula is used for computing the standard error. As we have seen this will be true when the serial correlation coefficients have the same sign, and in practice this aspect could be examined in the early stages of the investigation.

If the correlation coefficient is, on the other hand, found to be significantly large when independence is assumed, then it would be necessary to use the more precise formula. An alternative method of arriving at the values of ρ_s and $\bar{\rho}_s$ will be considered in the next section.

In the conclusions of his paper Bartlett (1935) raises the important point that if either of the parent series is serially independent then the standard significance tests are valid. This is apparent, since in either form of the precise formula if either of the serial correlation coefficients is zero, the standard error reduces to $n^{-\frac{1}{2}}$, the same as in the independent case.

4. SERIAL CORRELATIONS - DISTRIBUTION AND TESTS

Quenouille (1948) investigates the approximate distribution of the serial correlation coefficient defined by

$$r_{(1)} = \left[\sum_{i=1}^{n-1} x_i \; x_{i+1}\right] \cdot \left[\sum_{i=1}^{n-1} x_i^2 \quad \sum_{i=1}^{n-1} x_{i+1}^2\right]^{-\frac{1}{2}}$$

and shows that it is the same as that of the ordinary correlation coefficient which is based on (n+3) pairs of observations. In this context x_{n+i} is taken as x_i, and the result is adequate for n equal to or greater than 10.

It was shown in Section 1 above that r^2 is a $\beta_1\left(\frac{1}{2}, \frac{1}{2}(n-2)\right)$ variate and since the range of r is from -1 to +1 then

$$dp = \frac{(1-r^2)^{\frac{1}{2}(n-4)}}{B\left(\frac{1}{2}, \frac{1}{2}(n-2)\right)} dr$$

Then replacing n by n+3 and writing

$$\Gamma\left(\tfrac{1}{2}\right) . \Gamma\left(\tfrac{n+1}{2}\right) \Big/ \Gamma\left(\tfrac{n+2}{2}\right) \text{ for } B\left(\tfrac{1}{2}, \tfrac{1}{2}(n+1)\right) \text{ the distribution}$$

becomes

$$dp = \frac{(1-r^2)^{\frac{1}{2}(n-1)}\,\Gamma(n/2 + 1)}{\Gamma\left(\frac{1}{2}\right) \cdot \Gamma(n/2 + 1/2)} \; dr$$

and this is the form in which Quenouille writes it. If the x_i are connected by a linear Markoff scheme, i.e. if $x_{i+1} = \rho x_i + \varepsilon_{i+1}$ where the ε_i are independently and normally distributed about zero, then the distribution of the first serial correlation is given approximately by a modification of the distribution set out above. It is

$$dp = h(r)\; dr = \frac{(1-r^2)^{\frac{1}{2}(n-1)}\,\Gamma\left(\frac{n}{2} + 1\right)}{\Gamma\left(\frac{1}{2}\right) \cdot \Gamma\left(\frac{n}{2} + \frac{1}{2}\right) \cdot (1-2\rho r + \rho^2)^{n/2}}$$

The moments of this distribution are obtained by hyperbolic transformation, and the transformed value z (where $r = \tanh z$) is distributed approximately normally with the mean and variance both functions of ρ.

The theory is applied to several examples, one of which consists of serial correlations of daily pressure data for two six-monthly periods for the years 1930-1936. The original calculations are due to Walker (1946). The serial correlations have been divided into monthly groups and sample values of S^2 and theoretical values of σ^2 have been computed, the latter by means of the transformation mentioned above. The chi square distribution with 5 d.f. is then used to test the probability of occurrence of the estimated values.

Quenouille, in this paper, concludes that the approximate normal theory provides a satisfactory test for serial correlation coefficients provided n is sufficiently large.

This method of estimation of ρ enables a test to be made of the correlation between corresponding terms of two serially correlated series. Thus with the notation of Section 3 above, if ρ_1 is the correlation between successive terms of one series, and ρ_2 is the correlation between successive terms of another, then the correlation between corresponding terms is tested with
$n' = n\,(1-\rho_1\rho_2)\cdot(1+\rho_1\rho_2)^{-1}$ d.f.
This means the test is carried out as if there were n' pairs instead of n, and as mentioned earlier, if ρ_1 and ρ_2 have the same sign then n' is less than n and so the SE. of the correlation coefficient will be greater.

In a later paper, Quenouille (1958) considers the problem of comparing correlation structures within two time series, or within two lengths of the same series. That is, given r_s and r_s' ($s = 1, 2$ —— p) estimated from runs of n and m observations respectively, it is required to test whether they are consistent with a single set of parameters ρ_s. The author points out that no simple exact test of significance exists, but he does suggest a test procedure which involves fitting an auto-regressive series to the two series taken together, calculating the serial correlations and the partial serial correlations, and then testing the differences of the partial serial correlations with the chi square distribution. The test is based on the fact that in large samples the partial serial correlations of an auto-regressive series are independently and normally distributed with asymptotic variances $(n-s)^{-1}$.

5. CONCLUSIONS

It has been shown that the correlation coefficient estimated from samples of two series provides an indication of the way in which the two series behave with respect to each other. It has also been demonstrated that as the value of these observed correlation coefficients depends on the numbers of pairs used in the calculations, these numbers should always be quoted when sample correlation coefficients are compared.

The standard tests for the significance of an observed correlation coefficient are valid only if there is no serial correlation present in at least one of the two series under consideration. In meteorological series, these conditions are usually not realised, and the detail and precision of the tests carried out will have to be decided by the particular problem on hand.

The first step in testing significance of correlation between meteorological series, could be to use a standard test, and if a non-significant value of r has occurred, the sample serial correlations could be computed to check that they have the same sign. If this condition is also fulfilled, it could be assumed that the true correlation between the two series if not significant.

If either or both of these conditions are not fulfilled, and detail and accuracy are required, the test of whether the true correlation between the series is significant could be carried out by computing the SE. (r) as $(n^{-1}\ (1+2\sum \rho_s\ \bar{\rho}_s))^{\frac{1}{2}}$ as shown in Section 3, where the ρ_s and $\bar{\rho}_s$ are calculated by, say the transformation method of Section 4. This method will also provide a means of testing the significance of the serial correlations, since their distributions are found in the calculations.

REFERENCES

Bartlett, M.S. (1935) Some Aspects of the Time Correlation Problem in Regard to Tests of Significance. Journ. R. Stats. Soc., 48 Pt. 3, 536-543.

Fisher, R.A. (1948) Statistical Methods for Research Workers. Oliver and Boyd.

Moran, P.A.P. (1947) Some Theorems on Time Series I, Biometrika 34, 281-291.

Moran, P.A.P. (1949) The Statistical Analysis of the Sunspot and Lynx Cycles. Journ. Animal Ecology 18 No. 1, 115-116.

Quenouille, M.H. (1948) Testing of Aerial Correlation Coefficients, Biometrika 35 Pt 3, 261-267

Quenouille, M.H. (1958) The Comparison of Correlations in Time - Series, Journ. R. Stats. Soc., 20 No. 1 (B) 158-164.

Walker, Sir G. (1946) On Periods and Symmetry Points in Pressure as Aids to Forecasting, Q.J.R. Met. Soc. 72, 265-283.

Weatherburn, C.E. (1947) Mathematical Statistics, Camb. Univ. Press.