THE REALITY OF REGULARITIES INDICATED IN SEQUENCES OF OBSERVATIONS

G. F. McEWEN

SCRIPPS INSTITUTION OF OCEANOGRAPHY, LA JOLLA, CALIFORNIA

SUPPOSE THAT A SEQUENCE of observations exhibits a striking regularity; for example, suppose that the values arrange themselves in their order of magnitude, either increasing or decreasing, or that they indicate a maximum or a minimum. Is the indicated regularity real, is it true of the sequence or statistical set sampled, or is it the result of sampling from a random sequence? That is if the process were repeated, is there reason to believe that approximately the same general result would recur? The more improbable it is that the indicated regularity could have arisen as a result of random sampling, the greater is the justification for regarding the indications of the sample as truly representative. Unless the probability of obtaining an indicated regularity is small, there is no evidence of the existence of an actual regularity of the indicated type.

First, assuming that only the sequence of individual numerical quantities is available, we may apply various tests based upon characteristics of a random sequence (see Besson, 1920; Clough, 1921, 1924; Crum, 1921; Woolard, 1925; Working, 1934; Kermack and McKendrick, 1936–37; as well as a series of later and more exhaustive contributions). For example, the number of maxima in a sequence of unrelated numbers is one-third of the number of terms. The departure of a sequence in any characteristic from that deduced for a random sequence suggests some systematic influence, and the evidence in favor of such an influence depends upon the extent of the departure and the number of terms. In general, such laws of chance sequences do not serve as criteria for proving the presence of some systematic influence, unless there are a large number of terms. Moreover, the results of attempting to determine the probability of obtaining a short sequence of terms having a striking appearance of regularity tend to be rather misleading.

Second, form a sequence of averages of groups of individual observations selected from a given sequence in a known systematic way. For example, a sequence might be repeated any number of times. Then averages of corresponding terms would form a composite sequence to be tested. The statistical significance of such a sequence of averages may be determined by comparing the variance of the individual observations computed directly with that derived from the variances of the averages. This analysis of variance principle can be extended to the general case in which values of the dependent variable are related to each of a number of correlated independent variables. In this problem of multiple curvilinear correlation a sequence of averages of the dependent variable can be computed corresponding to each independent variable, and corrected to constant values of the other independent variables. A method is devised for testing the statistical significance of each such sequence of averages, as well as the composite significance of all of the sequences.

Reality of an Indicated Regularity of Averages Corresponding to Values of One Independent Variable

Frequently the values in a sequence are averages of measurements or numbers grouped in some systematic way. Statistical methods have been derived for calculating probabilities in such cases and depend upon the variability between the averages and within the groups. For practical purposes readily applicable methods are needed for determining such probabilities, and approximate estimates are acceptable. Investigations of the special case where the indicated regularities of a sequence are periodic are presented in numerous publications following the pioneer work done by Schuster and Buys-Ballot before 1900 (see Alter, 1927, 1933; Chree, 1924; Cox, 1923, 1924; Dodd, 1927; Kuznets, 1919; Powell, 1930; Stumpff, 1925, 1937; Walker, 1930).

The standard deviation of the group means multiplied by the square root of the number of observations used in computing a mean is one estimate of the standard deviation of the entire series, which may also be computed directly. Let

m = number of columns or groups
n = number of entries per column
$\bar{y}_s$ = a group mean
$\bar{y}$ = the grand mean

Then, if the entire arrangement is assumed to be accidental,

$$\sqrt{n}\,\sigma_n = \sqrt{n}\left[\sqrt{\frac{\sum(\bar{y}_s - \bar{y})^2}{m}}\right]$$

should not differ from

$$\sigma_0 = \sqrt{\frac{\sum(y - \bar{y})^2}{mn}}$$

by more than would be expected from sampling errors. Using this principle, Cox (1924) employed the following criterion of significance, often used by engineers and meteorologists. If the accidental error of σ_0 is assumed to be relatively small, the standard error of the ratio $\frac{\sqrt{n}\sigma_n}{\sigma_0}$ should be proportional to that of σ_n. He used the approximation $\sigma_n/\sqrt{2m}$ for this standard error. Accordingly, he compared the difference $\left[\frac{\sqrt{n}\sigma_n}{\sigma_0} - 1\right]$, which for a chance arrangement should on the average equal zero, with a standard error, $\frac{\sqrt{n}\,\sigma_n}{\sigma_0\sqrt{2m}}$. According to usual practice, a difference exceeding two or three times this standard error is evidence of some systematic relation between the averages. He used such a criterion for the existence of a periodic law of rainfall. This method is based upon the assumptions of the theory of large samples and may therefore often be inapplicable. Moreover, it would be better to compare $\sqrt{n}\,\sigma_n$ with the standard deviation of individual results from their sample means rather than with σ_0, which contains both sources of variability. This method may be improved by means of correction factors depending upon the size of the sample (E. S. Pearson, 1935). Thus for σ_n write $1/b_m\,(s)$, where s denotes the estimated standard deviation based upon m averages and, according to Pearson's table 13, $1/b_m$ has the following values:

$m =$	2	3	4	5	6	7	8	9	10
$1/b_m =$	1.772	1.382	1.253	1.189	1.151	1.126	1.108	1.094	1.084

$m =$	11	12	13	14	15	16	20	25	30
$1/b_m =$	1.075	1.068	1.063	1.058	1.054	1.050	1.040	1.031	1.026

In general the foregoing method, though widely used, is too uncertain and should be discarded.

The statistical significance of indicated regularities will now be considered with the aid of certain modern statistical principles whose detailed explanation can be found in the references cited. The probability that the variation between the averages of each column is due to chance may be estimated by means of the method of analysis of variance (Fisher, 1932; Tippett, 1931). An outline of the procedure follows:

m = number of columns
n_s = number of entries in column s
N = total number of entries = $\sum n_s$
y = any individual observation
$\bar{y}$ = grand mean of all the observations
a = a convenient assumed value of $\bar{y}$ to use in making computations
h = $(\bar{y} - a)$
$\bar{y}_s$ = mean of observed values in column s

$$V_s = \text{mean variance between columns} = \frac{\sum_1^m n_s(\bar{y}_s - a)^2 - Nh^2}{(m-1) = n_1}$$

$$V_r = \text{residual variance} = \frac{\sum(y-a)^2 - \sum_1^m n_s(\bar{y}_s - a)^2}{(N-m) = n_2}$$

$$Z = \log_e \sqrt{V_s/V_r}$$

$$\left.\begin{array}{l} n_1 = m - 1 \\ n_2 = N - m \end{array}\right\} = \text{degrees of freedom}$$

From Fisher (1932), table VI; Yule and Kendall (1937), appendix tables 6A, 6B, 6C; Fisher and Yates (1938), table V, read off the probability corresponding to Z and the degrees of freedom n_1 and n_2 of getting a variability between column means as great as that observed, or greater, corresponding to the variability of the individual values within the groups or columns.

Values for Z for only the 20 %, 5 %, 1 %, and 0.1 % points have been published, but results can be calculated approximately as follows for other points: If n_1 and n_2 are large or if they are approximately equal, the distribution of Z tends to approach normality, the standard deviation being (Tippett, 1931)

$$\sigma_z = \sqrt{\frac{1}{2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}.$$

Accordingly, the probability is the complement of that in Yule and Kendall (1937), appendix table 2, or in Pearson (1930) table II, corresponding to

$$T = Z \div \sqrt{\frac{1}{2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}.$$

An approximate method of interpolation or extrapolation, adapted to cases in which neither n_1 nor n_2 need be large, is to read off theoretical values of Z corresponding to the 20 %, 5 %, 1 %, 0.1 % points from table V, Fisher and Yates (1938). Then calculate values of T as follows:

T = 0.842 (observed Z) / (theoretical Z for 20 % point)
T = 1.645 (observed Z) / (theoretical Z for 5 % point)
T = 2.326 (observed Z) / (theoretical Z for 1 % point)
T = 3.090 (observed Z) / (theoretical Z for 0.1 % point)

and read off the probability corresponding to the normal law (Yule and Kendall, 1937, appendix table 2; or Pearson, 1930, table II). The multipliers, 0.842, 1.645, 2.326, and 3.090, correspond to the correct probabilities if the observed Z equals the theoretical Z, and it is assumed that T varies in proportion to the observed Z as in the previous formula. This method gives a closer approximation as the values of n_1 and n_2 increase.

Another method of interpolation that may also provide values extrapolated somewhat beyond the tabular limits depends upon the approximate relation

$$Z = A - B \log P,$$

found empirically from the tables, where A and B are constants for given values of n_1 and n_2. Accordingly, if values of Z corresponding to any values of n_1 and n_2 are plotted on the uniform scale of semi-logarithmic paper and values of the probability P are plotted on the logarithmic scale, the probability corresponding to the observed Z may be read from a straight line between appropriate points. For larger values of n_1 and n_2 the approximate value of the probability P may be calculated from Fisher's formulas, where the functions $A(P)$, $B(P)$, and λ may be plotted from values corresponding to P = 0.1 %, 1 %, 5 %, and 20 % (Fisher and Yates, 1938, table V; and the supplementary results found by Cochran, 1940). Results equivalent to Fisher (1932), table VI, are given by Snedecor (1934), table XXXV, directly in terms of the ratio $F = V_s/V_r$.

The foregoing methods of testing apparent regularities in series of observations are illustrated by numerical applications to two classes of natural phenomena, seasonal rainfall and sunspot numbers. First, consider the data, as given in table 1, on the seasonal rainfall at San Diego, California, for the 50-year interval from 1884 to 1933 inclusive, tabulated to investigate the existence of a 5-year cycle.

TABLE 1

SEASONAL RAINFALL AT SAN DIEGO, CALIFORNIA

Years	*1*	*2*	*3*	*4*	*5*
1884–1888........	8.7	17.0	8.3	9.8	11.0
1889–1893........	15.0	10.5	8.7	9.3	5.0
1894–1898........	11.9	6.2	11.8	5.0	5.2
1899–1903........	6.0	10.4	6.2	11.8	4.4
1904–1908........	14.3	14.7	10.6	8.5	10.2
1909–1913........	9.8	12.0	10.7	6.0	9.8
1914–1918........	14.4	12.5	10.1	8.0	8.7
1919–1923........	8.9	7.1	18.6	6.4	5.7
1924–1928........	5.8	15.7	14.7	8.7	7.1
1929–1933........	10.7	10.8	13.2	10.6	4.3
Averages ($\bar{y}_s$)...	10.55	11.69	11.29	8.41	7.14

There is an indicated increase from the first average to the second, followed by a regular decrease from the second to the last, or fifth average.

Now apply the analysis of variance method (see above, p. 231), where

$$\begin{aligned}
m &= 5; n_s = 10; N = 50; \bar{y} = 9.816; n_1 = 5 - 1 = 4; n_2 = 50 - 5 = 45\\
\bar{y}_s &= 10.55, 11.69, 11.29, 8.41, 7.14\\
V_s &= \sum 10(\bar{y}_s - 9.816)^2 \div (5 - 1) = 10 \times 15.357 \div 4 = 38.4\\
V_r &= [\sum(y-10.55)^2 + \sum(y-11.69)^2 \div \sum(y-11.29)^2 + \sum(y-8.41)^2\\
&+ \sum(y - 7.14)^2] \div (50 - 5) = [100.48 + 109.13 + 112.71 + 40.70\\
&+ 59.71] \div 45 = 422.73 \div 45 = 9.4\\
Z &= \log_e \sqrt{38.4}/\sqrt{9.4} = 0.703
\end{aligned}$$

From Fisher and Yates (1938), table V, the values of Z for the 20 %, 5 %, 1 %, and 0.1 % points are respectively 0.22, 0.48, 0.67, and 0.90. Therefore, since 0.703, the observed value of Z, is somewhat greater than 0.67, the probability is less than 0.01. The approximate extrapolated value of the probability corresponding to the median of the last three values of T in

$$\begin{aligned}
0.84 \times 0.703 \div 0.22 &= 2.68\\
1.65 \times 0.703 \div 0.48 &= 2.42\\
2.33 \times 0.703 \div 0.67 &= 2.44\\
3.10 \times 0.703 \div 0.90 &= 2.43
\end{aligned}$$

is 0.008, according to Yule and Kendall (1937), appendix table 2, or Pearson (1930), table II, and may be regarded as small enough to establish the reality of the indicated cycle.

The second illustration pertains to sunspot numbers arranged with reference to a trial cycle of length, 11 years, excluding extremes in which the sunspot number exceeded 99. Consider the data corresponding to the early interval, 1749 to 1826 inclusive, presented in table 2.

TABLE 2

SUNSPOT NUMBERS FROM 1749 TO 1826*

Years	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*
1749–1759..........	81.0	83.0	48.0	48.0	31.0	12.0	10.0	10.0	32.0	48.0	54.0
1794–1804..........	41.0	21.0	16.0	6.0	4.0	7.0	15.0	34.0	45.0	43.0	48.0
1805–1815..........	42.0	28.0	10.0	8.0	3.0	0.0	1.0	5.0	12.0	14.0	35.0
1816–1826..........	46.0	41.0	30.0	24.0	16.0	7.0	4.0	2.0	9.0	17.0	36.0
Averages ($\bar{y}_s$).....	52.5	43.2	26.0	21.5	13.5	6.5	7.5	12.7	24.5	30.5	43.2

* Exclusive of groups in which sunspot numbers exceed 99.

There is an indicated regular decrease from the first average through the sixth, followed by a regular increase to the last, or eleventh average.

Now apply the analysis of variance method, where

$$
\begin{aligned}
m &= 11;\ n_s=4;\ N=44;\ \bar{y}=25.59;\ n_1=11-1=10;\ n_2=44-11=33\\
\bar{y}_s &= 52.5, 43.2, 26.0, 21.5, 13.5, 6.5, 7.5, 12.7, 24.5, 30.5, 43.2\\
V_s &= \sum 4(\bar{y}_s - 25.59)^2 \div (11-1) = 956.32\\
V_r &= [\sum(y-52.5)^2 + \sum(y-43.2)^2 + \cdots] \div 33 = [1097.0 + 2312.76\\
&\quad + \cdots] \div 33 = 265.0\\
\sqrt{V_r} &= \sqrt{265.0} = 16.27\\
Z &= \log_e \sqrt{956.32}/\sqrt{265.0} = \log_e 1.898 = 0.6408
\end{aligned}
$$

For $n_1 = 10$ and $n_2 = 33$, the values of Z corresponding to the 20 %, 5 %, 1 %, and 0.1 % points are respectively 0.19, 0.38, 0.54, and 0.71 (Fisher and Yates, 1938, table V). Thus the probability is less than 0.01. The approximate extrapolated values of T are respectively

$$
\begin{aligned}
0.84 \times 0.6408 \div 0.19 &= 2.83\\
1.65 \times 0.6408 \div 0.38 &= 2.74\\
2.33 \times 0.6408 \div 0.54 &= 2.75\\
3.10 \times 0.6408 \div 0.71 &= 2.75
\end{aligned}
$$

and corresponding to the value, 2.75, the probability for n = infinity has the value 0.003, according to Yule and Kendall (1937), appendix table 2, or Pearson (1930), table II, which is so small that the evidence is conclusive in favor of the reality of the cycle.

In general such calculations of the probability corresponding to trial cycles of various lengths can be used as criteria in selecting the best estimate of cycle length.

Moreover, the same procedure, which was illustrated by cycle problems, can be applied to the combination of any number of sequences in which the observations in each sequence are tabulated with reference to the values of any independent variable, arranged in its order of magnitude.

Reality of a Regularity Indicated by Series of Group Averages Defining Multiple Regression

More generally, the dependent variable is a function of a number of independent variables which may be mutually correlated. The net relation of the dependent variable to any one of the independent variables considered can be presented by a series of averages of the dependent variable, corrected to constant values of the other independent variables and arranged according to increasing values of the selected independent one. For example, McEwen's method of successive approximations to group averages (McEwen and Michael, 1919; McEwen, 1920) is one way of making such a computation of multiple curvilinear regression without first assuming the forms of the functions.

An outline of the procedure for estimating the significance of the results obtained by this method follows:

M = total number of groups or columns to be averaged corresponding to all of the independent variables
N = number of values of the dependent variable
I = number of independent variables
y = any observed value of the dependent variable

$$\left.\begin{array}{l} {}_1\bar{y}_s \\ {}_2\bar{y}_s \\ \cdot \\ \cdot \\ \cdot \\ {}_I\bar{y}_s \end{array}\right\} =$$ corrected group averages of the dependent variable corresponding respectively to the 1st, 2nd, $\cdot\cdot\cdot$, Ith independent variable

${}_1\bar{y}$ = weighted average of values of ${}_1\bar{y}_s$
${}_2\bar{y}$ = weighted average of values of ${}_2\bar{y}_s$
$\cdot \qquad \cdot$
$\cdot \qquad \cdot$
$\cdot \qquad \cdot$
${}_I\bar{y}$ = weighted average of values of ${}_I\bar{y}_s$
n_s = number of entries in column s
V_s = overall variance between columns =

$$\frac{\sum n_s({}_1\bar{y}_s - {}_1\bar{y})^2 + \sum n_s({}_2\bar{y}_s - \bar{y})^2 + \cdot\cdot\cdot + \sum n_s({}_I\bar{y}_s - {}_I\bar{y})^2}{M - I = n_1}$$

iy' = any value of y corrected to constant values of all but the ith independent variables

$$V_r = \text{residual variance} = \frac{\sum_s \sum (iy' - iy_s)^2}{N + (I - 1) - M = n_2},$$

where i denotes any one of the independent variables. The numerator is also equivalent to $\sum(y - Y)^2$, where Y equals the computed value of the dependent variable. The significance is found as before from $Z = \log_e\sqrt{V_s/V_r}$.

Among the group averages of the independent variables, one, usually a central average, is selected for each independent variable. Then the group averages of the dependent variable corresponding to those of any independent variable are corrected to the selected constant averages of the other independent variables. Thus, one average of the dependent variables of each sequence will be the same for each independent variable. Accordingly, in computing V_r, the degrees of freedom equal

$$n_2 = N - [M - (I - 1)] = N + (I - 1) - M,$$

since this constant average should be counted only once. In the more elaborate treatment where a linear regression is computed in R of the groups, the foregoing value of n_2 would have to be reduced by R, and the averages in the numerator of V_r would be replaced by the corresponding linear regression function. Also, if a linear interpolation is computed for any number of intervals between successive groups, the foregoing value of n_2 should be reduced by the number of such intervals. At the most this number would be $(M - \mathrm{I})$. Finally, if it is desired to test the significance of the net relation to the ith independent variable, the value of the variance between columns would be

$$_iV_s = \frac{\sum n_s(i\bar{y}_s - i\bar{y})^2}{_im - 1},$$

where $_im$ equals the number of columns averaged to show the net relation to the ith independent variable, and $\sum_1^I {}_im = M$.

The significance of a computed value Y can be tested with the aid of "Student's" tables (1925), or table I, (pp. 118-119) of Pearson and Wishart (1942), where $n = n_2$ equals the number of degrees of freedom. The complement of the tabular entry is the probability that the average of a sample of size $n + 1$ will differ from the population average by more than t times the standard deviation of the average of the sample. Moreover, we may regard the complement of this tabular entry as the probability that a single observation in a sample of size $n + 1$ will differ from the population average by more than t times the standard deviation of the sample. Accordingly, we may use the results of the theory of small samples in estimating the significance of the computed value Y as follows: For $t = \alpha/\sqrt{V_r}$, and $n = n_2$ in the tables, the complement of the entry equals the probability that a computed value will exceed the observed value by more than $\alpha = t\sqrt{V_r}$.

REFERENCES

ALTER, D.

1927. "A group or correlation periodogram, with application to the rainfall of the British Isles," *Monthly Weather Review*, vol. 55, pp. 263-266.

1933. "An extremely simple method of periodogram analysis," *Proc. Natl. Acad. Sci.*, vol. 19, pp. 335-339.

BARTELS, J.

March 1935. "Random fluctuations, persistence, and quasi-persistence in geophysical and cosmical periodicities," *Terrestrial Magnetism and Atmospheric Electricity*, vol. 40, pp. 1-60.

BESSON, LOUIS

1920. "On the comparison of meteorological data with results of chance" (translated from the French, and abridged, by Edgar W. Woolard), *Monthly Weather Review*, vol. 48, pp. 89-94.

BEVERIDGE, W. H.

1922. "Wheat prices and rainfall in western Europe," *Jour. Roy. Stat. Soc.*, vol. 85, pp. 412-478.

CAMP, B. H.

1931. *The Mathematical Part of Elementary Statistics*. D. C. Heath & Co.

CHREE, C.

1924. "Periodicities, solar and meteorological," *Jour. Roy. Meteorol. Soc.*, vol. 50, pp. 87-97.

CLOUGH, H. W.

1921. "A statistical comparison of meteorological data with data of random occurrence *Monthly Weather Review*, vol. 49, pp. 124-132.

1924. "A systematically varying period with an average length of 28 months in weather and solar phenomena," *ibid.*, vol. 52, pp. 421-441.

COCHRAN, W. G.

1940. "Note on approximate formula for the significance levels of Z," *Annals of Math. Stat.*, vol. 11, pp. 93-95.

COOLIDGE, J. L.

1925. *An Introduction to Mathematical Probability*. Oxford University Press.

COX, JOEL B.

Oct. 1923. "Periodic fluctuations of rainfall in Hawaii," *Proc. Amer. Soc. Civil Eng.*, vol. 69, pp. 1699-1721.

1924. "Periodic fluctuations of rainfall in Hawaii," *Trans. Amer. Soc. Civil Eng.*, vol. 87, pp. 461-491.

CRUM, W. L.

Dec. 1921. "A measure of dispersion for ordered series," *Amer. Stat. Assoc. Quart. Publ.*, vol. 17, pp. 969-975.

DODD, E. L.

1927. "The probability law for the intensity of a trial period, with data subject to the Gaussian law," *Bull. Amer. Math. Soc.*, vol. 33, pp. 681-684.

EZEKIEL, M.

1941. *Methods of Correlation Analysis*. John Wiley & Sons, New York.

FISHER, R. A.

1932. "Statistical methods for research workers," *Biological Monographs and Manuals*, 4th ed. London.

FISHER, R. A. and F. YATES

1938. *Statistical Tables for Biological, Agricultural and Medical Research*. Oliver & Boyd, London.

FOWLE, F. E.

1920. *Smithsonian Physical Tables*. Smithsonian Institution, Washington, D.C.

KERMACK, W. O., and A. G. MCKENDRICK

1937. "Tests for randomness in a series of numerical observations," *Proc. Roy. Soc. Edinburgh*, vol. 57 (1936-37), pp. 228-240.

"Some distributions associated with a randomly arranged set of numbers," *ibid.*, pp. 332-376.

KUZNETS, SIMON

1929. "Random events and cyclical oscillations," *Jour. Amer. Stat. Assoc.*, vol. 24, pp. 258-275.

MARVIN, C. F.

1921. "Theory and use of the periodocrite," *Monthly Weather Review*, vol. 49, pp. 115-133.

MCEWEN, G. F.

1920. "The minimum temperature, a function of the dew point and humidity, at 5 p.m. of the preceding day; method of determining this function by successive approximations to group averages," *Monthly Weather Review*, Suppl. no. 16, pp. 64-69.

MCEWEN, G. F., and E. L. MICHAEL

1919. "The functional relation of one variable to each of a number of correlated variables determined by a method of successive approximations to group averages," *Proc. Amer. Acad. Arts and Sci.*, vol. 55, pp. 89-133.

PEARSON, E. S.

1935. "The application of statistical methods to industrial standardization and quality control," *Report, British Standards Institution* (London), vol. 600, pp. 1-161.

PEARSON, E. S., and JOHN WISHART

1942. *"Student's" Collected Papers*. Cambridge. University Press.

PEARSON, K.

1930. *Tables for Statisticians and Biometricians, Part 1*. 3rd ed. Cambridge University Press, England.

POWELL, RALPH W.

May "Successive integration as a method for finding long period cycles,"
1930. Annals of Math. Stat. vol. 1, pp. 123-136.

SNEDECOR, GEORGE W.

1934. *Calculation and Interpretation of Analysis of Variance and Covariance*. Ames, Iowa, Collegiate Press, Inc.

"STUDENT"

1925. "New tables for testing the significance of observations," *Metron*, vol. 5 (1925), p. 105.

STUMPFF, K.

1925. "Analyse Periodischer Vorgänge," Zeitsch f. angew. Geophys., Bd. 1.

1937. *Grunlagen und Methoden der Periodenforschung*. Springer, Berlin.

TIPPETT, L. H. C.

1931. *The Methods of Statistics*. Williams & Norgate, Ltd., London.

WALKER, SIR GILBERT T.

1930. "On periodicity—criteria for reality," *Mem. Roy. Meteorol. Soc.*, vol. 3, no. 25, pp. 97-101.

WHITTAKER, E. T., and G. ROBINSON

1932. *The Calculus of Observation*. Blackie & Son, Ltd., London and Glasgow.

WHITWORTH, W. A.

1927. *Choice and Chance*. 5th ed., reprint. Stechert, New York.

WOOLARD, Edgar W.

1925. "On the mean variability in random series," *Monthly Weather Review*, vol. 53, pp. 107-111.

WORKING, HOLBROOK

1934. "A random-difference series for use in the analysis of time series," *Jour. Amer. Stat. Assoc.*, vol. 24, pp. 11-24.

YULE, G. UDNY, and M. G. KENDALL

1937. *An Introduction to the Theory of Statistics*. 11th ed. Charles Griffin & Co., Ltd., London.