JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1982, 15, 423-429 NUMBER 3 (FALL 1982)

A SIMPLIFIED TIME-SERIES ANALYSIS FOR EVALUATING TREATMENT INTERVENTIONS

WARREN W. TRYON

FORDHAM UNIVERSITY

Time-series analysis procedures for analyzing behavioral data are receiving increasing support. However, several authorities strongly recommend using at least 50–100 points per experimental phase. A complex mathematical model must then be empirically developed using computer programs to extract serial dependency from the data before the effects of treatment interventions can be evaluated. The present discussion provides a simple method of evaluating intervention effects that can be used with as few as 8 points per experimental phase. The calculations are easy enough to do by hand.

DESCRIPTORS: time-series analysis, statistics, statistical inference

Time-series analysis is a quantitative method for assisting the subjective art of data interpretation. Several studies (e.g., Gottman & Glass, 1978; Jones, Weinrott, & Vaught, 1978) have provided empirical demonstrations that visual and statistical evaluations of typical applied behavior analytic data sets often differ. Although it is not universally agreed that statistically based judgments are better than visual judgments (e.g., Baer, 1977), many authors (Hartmann, Gottman, Jones, Gardner, Kazdin, & Vaught, 1980; Jones, Vaught, & Weinrott, 1977; McCain & McCleary, 1979) have argued that it is desirable to use time-series statistics to analyze behavioral data. More confidence can be placed in data interpretations when statistical and visual analyses agree than when they disagree.

The particular time-series analysis most often suggested (e.g., Glass, Willson, & Gottman, 1975) is based on an auto-regressive integrated moving average. This procedure involves the empirical construction of a complex mathematical model that is subsequently validated against the very data from which it was constructed. The model is used to extract serial dependency from the data only after all the criteria of model construction have been met, thereby leaving an uncorrelated time series. Standard inferential statistics are used to evaluate changes in the mean level and slope of the time series due to the intervention.

A major limitation of the auto-regressive integrated moving average approach is that many data points are required for adequate model development. Hartmann et al. (1980) cited several authorities who recommended collecting at least 50 to 100 data points *per experimental phase* before attempting to use auto-regressive integrated moving average procedures. Less confidence exists in the empirically constructed model if fewer data points are used. Moreover, the power of the auto-regressive integrated moving average procedure is diminished as data are reduced, thus increasing the probability of falsely accepting the null hypothesis.

The purpose of the current discussion is to present a simple, yet elegant, method of time-series analysis that can be used on small data sets to evaluate the effects of treatment interventions. This approach can also be used to decide when responding has stabilized, i.e., when a new phase of the experiment might begin (cf. Killeen, 1978). The logic underlying the *C* statistic is the same as the logic underlying visual analysis; variability in successive data points is evaluated relative to changes in slope from one phase of the experiment to another.

Reprint requests should be sent to Warren W. Tryon, Department of Psychology, Fordham University, Bronx, New York 10458.

THE C STATISTIC

vonNeumann, Kent, Bellinson, and Hart (1941) described two orthogonal estimates of the variance of a time series. The first measure is the variance calculated as indicated in Equation 1.

$$S^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - X)^2 \qquad (1)$$

This variance of the time series increases in direct proportion to changes or trends in the mean value of the series. Consider the following data: 1, 2, 3, 4, 5. Their mean is 3 and their variance is 2.5. If this trend extends to include: 1, 2, 3, 4, 5, 6, 7, 8, 9; then their mean is 5.0 and their variance is 7.5. Hence, the presence of a trend increases both the mean and the variance. Said otherwise, the variance is inversely proportional to the stationarity of the series.

The second estimate of the variance of a time series is the Mean Square Successive Difference (*MSSD*) statistic. It is calculated as its name implies. The consecutive differences among data points are calculated, squared, and then averaged as indicated by Equation 2.

$$MSSD = D^2 = \sum_{i=1}^{N-1}\frac{(X_{i+1} - X_i)^2}{N-1} \qquad (2)$$

The *MSSD* or *D* squared statistic is independent of changes in the mean value of the time series, i.e., it is independent of the stationarity of the series. Reconsider the two brief data sets given above. The *MSSD* statistic equals 1.0 for both sets: integers 1-5 and integers 1-9. Notice that the continuing trend increased the mean squared deviation from the mean by a factor of 3 but did not alter the mean squared successive difference.

vonNeumann (1941) extensively discussed the distribution of the ratio of the *MSSD* to the variance. However, it was Young (1941) who developed this reasoning into the highly useful *C* statistic given by Equation 3.

$$C = 1 - \frac{\sum_{i=1}^{N-1}(X_i - X_{i+1})^2}{2\sum_{i=1}^{N}(X_i - X)^2} \qquad (3)$$

The numerator of the right-hand term is the sum of the $N - 1$ squared consecutive differences associated with the time series. The denominator of this same term is twice the sum of the N squared deviations of the time-series data points from their mean.

The standard error of the *C* statistic is easily calculated using Equation 4 and it depends entirely on the number of data points in the time series.

$$Sc = \sqrt{\frac{N+2}{(N-1)(N+1)}} \qquad (4)$$

Young (1941) has shown that the ratio of *C* to its standard error is the *Z* statistic

$$Z = \frac{C}{Sc} \qquad (5)$$

and is normally distributed for time series containing 25 or more values. Moreover, the deviation from normality is not marked even for time series containing just 8 values. Table 1 presents the 5% and 1% critical values for samples of size 8 to 25.

Characteristics of the C Statistic

Reference to Equation 3 will help illustrate the basic characteristics of the *C* statistic. The value of *C* will be zero when the sum of the squared deviations from the mean equals one-half the sum of the squared consecutive differences, because the denominator of the right-hand fraction is multiplied by 2 which makes it equal to the numerator and thus the right-hand fraction equals unity. Subtracting unit from one leaves zero. This situation is most likely to occur when the data hug the mean rather closely.

The sum of squared deviations from the mean increases more rapidly than does the sum of

Table 1

Critical values for testing the *C* statistic for selected sample sizes (*N*) at the .01 level of significance[a, b].

N	*1%*	*N*	*1%*
8	2.17	18	2.25
9	2.18	19	2.26
10	2.20	20	2.26
11	2.21	21	2.26
12	2.22	22	2.26
13	2.22	23	2.27
14	2.23	24	2.27
15	2.24	25	2.27
16	2.24	∞	2.33
17	2.25		

[a]Taken from Young (1941).

[b]The critical value for the .05 level of significance is 1.64 for all sample sizes above.

squared successive differences, given the presence of any type of trend or nonstationarity. This causes the right-hand fraction of the *C* statistic to become small, which makes the *C* statistic become large. The *C* statistic aids the investigator in evaluating how large the squared deviations from the mean are (which reflect the presence of all types of trends) relative to the sum of the squared consecutive differences (which are independent of all types of trends). The logic of this fraction is directly analogous to that of the *F* statistic.

The statistical significance of *C* is evaluated by dividing it by its standard error (cf. Equation 5). It should be noted that the standard error is entirely a function of sample size. This means that the standard error can be reduced to any value, and thus a significant *Z* can always be found given any value of *C*. Hence, the power of the test approaches infinity as the sample size approaches infinity. Entirely trivial effects can be found to the statistically significant if enough data points are collected. It should be noted that this predicament is generally true of all statistical analyses and is not a unique limitation of the *C* statistic.

Applying the C Statistic

The main logical question answered by the *C* statistic is whether or not the time series contains any trends, i.e., any systematic departures from random variation. An initial use of the *C* statistic is to evaluate the baseline data. Two outcomes are possible. Evidence of a trend will either be found or not. It is more desirable that the baseline data not contain any statistically significant trends because this allows a more powerful application of the *C* statistic by appending the first treatment series to the first baseline series and testing the ensemble or aggregate series with the *C* statistic. A significant result is evidence that the treatment series departs from the baseline series.

Two less powerful applications of the *C* statistic are available for use when the initial baseline is found to contain a trend. Both alternative procedures involve creating a comparison series and testing for a trend with the *C* statistics. The more powerful of these two alternative procedures involves calculating "difference scores from the trend in the previous phase" (Hayes, 1981, p. 201). Several methods are available for quantifying the trend in the previous phase. Standard regression techniques can be used to obtain a line of best fit. However, one or two atypical data points can severely affect both the slope and intercept values given small data sets. Velleman and Hoaglin (1981) describe how to fit a "resistant line" which passes through the medians of each third of the data. The slope and intercept values in the equation for the resistant line agree favorably with the corresponding values in the regression equation when no atypical data points are present. If a straight line does not adequately describe the trend in the previous phase, then a more complex equation is required. Perhaps a quadratic, polynomial, or trigonometric function, like a sine wave, characterizes the data more accurately. Daniel and Wood (1971) and Lewis (1960) are good sources of curve fitting procedures. The comparison series is obtained by subtracting the trend line values associated with the first baseline point from the first treatment point, then subtracting the trend line value associated with second baseline point from the second treatment

point, etc., until all baseline and/or treatment values have been exhausted. Often, more treatment points will exist than baseline points. Modest extrapolation of the first phase trend line can provide a basis for adding a few more points to the comparison series, thereby enhancing the power of the test. The comparison series is tested with the *C* statistic. A significant result is evidence that the difference between the trends in the two phases contains a trend or departure of some kind. A significant *C* statistic only establishes that change has occurred. It does not guarantee that the change was due to the variable manipulated by the experimenter; it could be due to changes in an uncontrolled collateral variable. As with visual analysis, it is the overall pattern of results relative to the design used that enables a determination that the independent variable is responsible for change.

The second of the less powerful applications of the *C* statistic is the easier to use. The comparison series is obtained directly by subtracting the first baseline value from the first treatment value, etc., until all baseline and/or treatment values have been exhausted. The *C* statistic is then calculated on this comparison series. A significant result indicates that the treatment phase departs from the trend set in baseline.

Both uses of a difference series (unlike the use of raw scores when baseline is stationary) share a common limitation. The *C* statistic will not be significant when the slopes of data points in the two experimental phases under consideration are equal even when one series has been shifted up or down dramatically relative to the other series. This is because the difference series will be constant, i.e., highly stationary.

The next use of the *C* statistic might be to determine when responding has stabilized during treatment. One criterion might be to continue data collection until 10 consecutive data points are obtained for which the *C* statistic was not significant (see Killeen, 1978 for other criteria). These data would then provide a period against which the subsequent phase could be assessed. This process would continue for each successive phase. Sometimes responding may not meet the stability criteria before treatment must be reinitiated or withdrawn. The less powerful alternative procedure for using the *C* statistic could then be used.

It may seem that if values associated with a linear trend in a previous phase can be subtracted from data in a subsequent experimental phase then these values could just as well be subtracted from a linear trend line associated with the subsequent phase. Such a procedure *always* gives artificial results. It can be shown that the value of the *C* statistic is always entirely a function of the sample size when values of one linear trend line are subtracted from values of another linear trend line. That is, the value of *C* associated with all sets of, say, 15 data points will be exactly the same regardless of the data used. A different value of *C* is associated with each value of *N*. This anomaly arises because the difference between two linear trend lines is itself a linear trend line where the differences between consecutive values is constant and equal to the slope of the comparison line. However, the sum of the squared deviations from the mean associated with values on this comparison line is also a function of its slope. The ratio of the sum of the squared consecutive differences is a constant fraction of the sum of the squared deviations from the mean of such values depending only on their number.

AN EMPIRICAL EXAMPLE

Tryon and Zager (Note 1) observed the frequency of "talking-out" behavior in a class of 15 mentally retarded children aged 9-11 yr. A "talk-out" was defined as all vocalizations not authorized by the teacher. Observations were made for 1 h in the morning and for 1 h during the afternoon, Monday through Friday, yielding 10 baseline data points during the first week. These data are displayed in Figure 1.

The first question was whether some trend existed in the baseline data. The actual data are presented in Table 2 where the *C* statistic has

Table 2

Example of the Use of the *C* Statistic in an A-B-A Experimental Design

	Score (*X*)	D^2	
First Baseline Phase	28	324	First Baseline Phase:
	46	49	$D^2 = 1112$
	39	36	$2\,SS(X) = 1324$
	45	441	$C = 1 - \frac{1112}{1324} = 0.160$
	24	16	$Sc = \sqrt{\frac{8}{9(11)}} = 0.284$
	20	225	$Z = \frac{0.160}{0.284} = 0.563$, n.s.
	35	4	
	37	1	
	36	16	
	40	256	
Group Tokens Phase	24	64	First Baseline Plus Group Tokens:
	16	441	$D^2 = 2762$
	37	64	$2\,SS(X) = 8227.0$
	45	729	$C = 1 - \frac{2762}{8227} = 0.664$
	18	1	$Sc = \sqrt{\frac{30}{31(33)}} = 0.171$
	19	1	$Z = \frac{0.664}{0.171} = 3.883$
	18	0	$p < .001$
	18	25	
	13	1	
	12	9	
	15	4	
	13	4	
	15	1	
	16	25	
	11	9	
	14	0	
	14	4	
	12	1	
	13	1	
	14	9	
	17	1	Last Week of Group Tokens Plus
	16	1	First Week of Second Baseline:
Second Baseline Phase	15	36	$D^2 = 353$
	21	25	$2\,SS(X) = 882$
	16	49	$C = 1 - \frac{353}{822} = 0.571$
	23	9	$Sc = \sqrt{\frac{18}{19(21)}} = 0.212$
	20	36	$Z = \frac{0.571}{0.212} = 2.693$
	26	0	$p < .01$
	26	16	
	22	49	
	15	81	
	24		

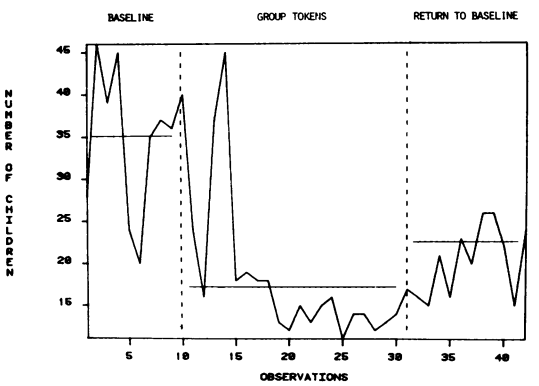

Fig. 1. The total number of children participating in all incidents of unauthorized talk-outs during baseline 1, group tokens, and baseline 2 phases.

been calculated. The value of $Z = .563$ is not statistically significant, indicating the absence of any substantial trend.

The next 11-day (22 observations) phase involved a group consequences procedure (cf. Barrish, Saunders, & Wolf, 1969; Herman & Tramontana, 1971; Packard, 1970; Schmidt & Ulrich, 1969). The teacher put a token in a glass jar on the teacher's desk at the end of every 5-min period during which no "talking-out" behavior occurred by any class member. Each student earned the number of tokens in the jar at the end of each period and tokens could be exchanged for edibles. The basic question at issue was whether the group tokens procedure had any demonstrable effect on "talking-out" behavior (see Figure 1). The data for this phase of the experiment were appended to the baseline data and tested for a trend. The resulting $Z = 3.883$, $p < .001$ confirmed the visual impression of a shift in the trend of the time series.

The next question concerned when responding under the group tokens procedure had stabilized. Inspection of the D2 column of Table 2 reveals large consecutive changes early in the intervention but leveling off shortly thereafter. An analysis of all 22 group tokens data points yielded a $Z = 2.468$, $p < .05$, confirming the visual evidence of a trend. The last 10 data points (1 wk of observation) were chosen to assess stability to compare this portion of the series with the first week of return to baseline. The value of $Z = .679$ was clearly not statistically significant, suggesting that this portion of the series was stable.

The group tokens procedure was discontinued for a 2-wk period which constituted a second baseline period. Table 2 contains the calculations associated with the last 10 data points from the intervention plus the first 10 data points from the second baseline. The resulting $Z = 2.693$, $p < .01$ indicated the presence of a trend.

Visual inspection of Figure 1 may suggest to some the presence of a trend occurring during the 2-wk second baseline period. The resulting value of $Z = .146$ is not statistically significant. This is consistent with observations (Gottman & Glass, 1978; Jones et al., 1978) that data analysis based on visual inspection and time-series analysis can disagree substantially.

DISCUSSION

The *C* statistic is a simple, yet elegant, method for quantitatively evaluating the presence of changes due to treatment interventions in serially dependent time-series data. It is an omnibus test for abrupt changes in the level of a time series as well as gradual changes in its slope. The major difference between the *C* statistic and the auto-regressive integrated moving average method is that the latter can test for abrupt changes in level separately from changes in slope while the former cannot. However, the *C* statistic can be used with much smaller data sets, does not require complex computer based model construction, and is easily calculated by hand.

The *C* statistic is best applied when responding has stabilized in the previous phase. Then the data from the subsequent phase can be appended to the previous phase and tested for any trends using the *C* statistic. Two alternate methods can be used when responding has not stabilized in the previous phase. The more powerful alternate method is to fit either a resistant or regression line to the data in the prior phase and then create a comparison series by subtract-

ing the trend line values associated with the previous phase from the data points in the subsequent phase. This comparison series is then tested for any trends using the *C* statistic. The less powerful alternate method involves subtracting corresponding data points in the previous phase from those in the subsequent phase to create the comparison series. This series is then tested for any trends using the *C* statistic as before. Both of the less powerful methods share the limitation that they cannot test for a change in level if there has been no change in slope.

This flexible and easily calculated time-series *C* statistic should be of use to investigators who did not previously have the resources to incorporate time-series designs into their research and/or clinical practice.

REFERENCE NOTE

1. Tryon, W. W., & Zager, K. Reduction of talking-out behavior in a class of mentally retarded children through group consequences. Unpublished manuscript, 1980. (Available from Department of Psychology, Fordham University, Bronx, New York 10458.)

REFERENCES

Baer, D. M. "Perhaps it would be better not to know everything." *Journal of Applied Behavior Analysis,* 1977, **10,** 167-172.

Barrish, H. B., Saunders, M., & Wolf, M. M. Good behavior game: Effects of individual contingencies for group consequences on disruptive behavior in a classroom. *Journal of Applied Behavior Analysis,* 1969, **2,** 119-124.

Daniel, C., & Wood, F. S. *Fitting equations to data: Computer analysis of multifactor data for scientists and engineers.* New York: McGraw-Hill, 1971.

Glass, G. V., Willson, V. L., & Gottman, J. M. *Design and analysis of time series experiments.* Boulder: Colorado Associated University Press, 1975.

Gottman, J. M., & Glass, G. V. Analysis of interrupted time-series experiments. In T. R. Kratochwill (Ed.), *Single subject research: Strategies for evaluating change.* New York: Academic Press, 1978.

Hartmann, D. P., Gottman, J. M., Jones, R. R., Gardner, W., Kazdin, A. E., & Vaught, R. Interrupted time-series analysis and its application to behavioral data. *Journal of Applied Behavior Analysis,* 1980, **13,** 543-559.

Hayes, S. C. Single case experimental design and empirical clinical practice. *Journal of Consulting and Clinical Psychology,* 1981, **49,** 193-211.

Herman, S. H., & Tramontana, J. Instructions and group versus individual reinforcement in modifying disruptive group behavior. *Journal of Applied Behavior Analysis,* 1971, **4,** 113-119.

Jones, R. R., Vaught, R. S., & Weinrott, M. Time-series analysis in operant research. *Journal of Applied Behavior Analysis,* 1977, **10,** 151-166.

Jones, R. R., Weinrott, M. R., & Vaught, R. S. Effects of serial dependency on the agreement between visual and statistical inference. *Journal of Applied Behavior Analysis,* 1978, **11,** 277-283.

Killeen, P. R. Stability criteria. *Journal of Experimental Analysis of Behavior,* 1978, **29,** 17-25.

Lewis, D. *Quantitative methods in psychology.* New York: McGraw-Hill, 1960.

McCain, L. J., & McCleary, R. The statistical analysis of the simple interrupted time-series quasi-experiment. In T. D. Cook & D. T. Campbell (Eds.), *Quasi-experimentation: Design and analysis issues for field settings.* Chicago: Rand-McNally, 1979.

Packard, R. G. The control of "classroom attention": A group contingency for complex behavior. *Journal of Applied Behavior Analysis,* 1970, **3,** 13-28.

Schmidt, G. W., & Ulrich, R. E. Effects of group contingent events upon classroom noise. *Journal of Applied Behavior Analysis,* 1969, **2,** 171-179.

Velleman, P. F., & Hoaglin, D. C. *Applications, basics, and computing of exploratory data analysis.* Boston: Duxbury Press, 1981.

vonNeumann, J. Distribution of the ratio of the mean square successive difference to the variance. *Annals of Mathematical Statistics,* 1941, **12,** 367-395.

vonNeumann, J., Kent, R. H., Bellinson, H. R., & Hart, B. I. The mean successive difference. *Annals of Mathematical Statistics,* 1941, **12,** 153-162.

Young, L. C. On randomness in ordered sequences. *Annals of Mathematical Statistics,* 1941, **12,** 293-300.

Received March 9, 1981
Final acceptance January 14, 1982