A Nonparametric Quasi-Experimental Technique for Evaluating Highway Traffic Safety Countermeasures

JESSIE C. FORTENBERRY, DAVID B. BROWN, and DAVID M. RHYNE

ABSTRACT

A nonparametric quasi-experimental time series analysis is presented as a method for circumventing many of the problems often encountered in applying statistical analysis to the evaluation of safety programs. It is shown how this method is applicable where classical experimental techniques are impossible to implement. A case study, which shows the application to an actual pedestrian program evaluation, is given to exemplify the technique.

Under ideal conditions, experiments are set up in such a manner that by varying certain factors under controlled conditions, observation can be made as to the effects of the factors involved. The manner in which these factors are combined, the sequence in which they are observed, and the conditions under which they are controlled constitute the experimental design. Unfortunately, not all situations that need evaluating lend themselves to the controls that are dictated by classical experimental design. It is not always possible to randomly assign treatment conditions or to control unknown factors. Under these circumstances, a "quasi-experimental" design may be considered.

A quasi-experimental design describes how observations can be made under constraints imposed by nonlaboratory environments. The primary concern is to isolate the effects of interest from outside influences. For evaluation studies, effects of interest are the impacts that a program has on the performance measures under consideration. The quasi-experimental design attempts to maximize the extent to which these effects can be measured without contamination from other effects such as historical trends, seasonal influences, parallel activities, or universal random effects. Under this design, the philosophy is simply to measure performance parameters in such a manner as to eliminate alternative explanations (other than program effects) for any change observed.

Conceptually, the situation consists of a time series of N observations as follows:

$$O_1 \quad O_2 \quad O_3 \quad O_4 \quad O_5 \quad O_6 \quad O_7 \quad \cdots \quad O_N$$

At some point in the time series, an intervention for a treatment area is observed at the point where the treatment is initiated.

$$O_1 \quad O_2 \quad O_3 \cdots O_M \mid O_{M+1} \quad O_{M+2} \quad O_{M+3} \cdots O_N$$

TREATMENT INITIATED

Here, N observations are taken with M observations occurring before the treatment is introduced and N - M observations occurring after the treatment. The first M observation can be considered as a base period and the N - M observations can be considered as constituting an operational period.

With such a time series of observations, analysis can be performed by measuring performance measures before and after a treatment condition is introduced. In a fairly stable environment, this type of analysis should be able to detect any effects that the treatment condition has on the performance measures. However, just because analysis of a time series of observations detects a change in performance measures that coincides with the initiation of a treatment program does not necessarily indicate that the change was due to the treatment. An evaluation must be able to first assess the effects of a treatment condition and then eliminate alternative explanations for such change.

One method of logically eliminating some alternative explanations for change is to use a control. This would consist of two time series of observations--one for treatment and one for control--as follows:

	Pre Treatment	Post Treatment
Control Area	$O \quad O_{12} \cdots O_{1m}$	$O_{1M+1} \; O_{1M+2} \cdots O_{1N}$
Treatment Area	$O_{21} \; O_{22} \cdots O_{2M}$	$O_{2M+1} \; O_{2M+2} \cdots O_{2N}$

This experimental design can handle a number of situations provided that adequate control areas and sufficient time periods are selected. Note that in this paper, an "ideal control area" is one that is identical in all aspects to the treatment area with the exception of treatment program. Obviously, this situation does not exist in the nonlaboratory environment. The next best thing is to select a control area that is similar to the treatment area in as many factors as possible. [For a more complete discussion of quasi-experiment design, see Caporaso and Roos (1).]

TECHNIQUE

The objective of this type of evaluation is to accurately compare the actual data for the operational period with what could reasonably be expected based on the data of the baseline period. The assumption is made that if the project intervention had not taken place, then the data for the operational period would be an undisturbed continuation of the data for the baseline period. A regression model is constructed as an effort to accurately project what can reasonably be expected for the operational period based on the data of the baseline period. These projected values are then compared with the actual data in evaluating the effect that the program intervention had on the data.

A new evaluation technique was developed that consists of a unique combination of two well-known evaluation procedures, linear regression and the Wilcoxon Test. It was developed from the concepts of the quasi-experimental design and the need for comparing a time series of data following program initiation to a time series of data that would have been expected had a program not occurred. It also involves the use of a control area to prevent uncontrollable factors from giving a false indication of program effectiveness.

One of the assumptions of regression models is that variability of the data is symmetrically distributed about the line of regression of y and x. If this assumption holds, then it is equally likely that the data will fall above the regression line as below the regression line. Now, if regression models were constructed for both a treatment area and a control area, then it also would be equally likely that as many of the data points for the control area will fall below the regression line as for the treatment area. Not only are the data points equally likely to fall above as below the regression line, but they should also be equally distributed about the regression lines in terms of magnitude when adjusted by their standard error of estimate. In other words, it is equally likely that x number of data points will fall +2 or more standard-error-of-estimate points below the regression line for the control area as for the treatment area.

The fact that the linear regression for one area may account for a larger proportion (R^2) of the variance than the linear regression for the other area does not present a problem for the technique. Even if the accounted variance for one area is smaller than for the other area, the remaining variability of the data is symmetrically distributed about the line of regression of y on x. The magnitude of this variability is then adjusted by its standard error of estimate. A large symmetrical variability about the regression line does not affect the test statistic. It is only when a set of data becomes unsymmetrical about the regression curve that the test statistic is affected. In other words, the test statistic is affected whenever there is a shift in the general trend in the data for one area and there is no corresponding shift in the general trend for the other area.

To evaluate a treatment area, the deviations (actual data minus forecasted data) from both the treatment area and the control area are converted to Z values (deviation-standard error of estimate). The data then consist of n' paired observations (Z_{T1}, Z_{C1}), (Z_{T2}, Z_{C2}), ..., (Z_{Tn}, Z_{Cn}), where Z_{Ti} is the Z value for the treatment area for the ith month and Z_{Ci} is the Z value for the control area for the ith month. The absolute differences (without regard to sign), which can be represented as

$$Di = |Z_{Ci} - Z_{Ti}|;\ i = 1, 2, \ldots, n' \qquad (1)$$

are then computed for each of n' pairs (Z_{Ti}, Z_{Ci}). The differences (Di's) are ranked from 1 to n' according to their absolute values.

If the common median of the Di's is denoted by $D_{.50}$ and the test is one-tailed, then the hypotheses may be stated as

$$H_0 : D_{.50} \leq 0$$

$$H_1 : D_{.50} > 0$$

The alternative hypothesis may be stated in words as "The values of Z_{Ti}'s tend to be smaller than the values of Z_{Ci}'s." The test statistic (T) is the sum of the ranks of the positive Di's. Large values of T indicate that H_0 is false, so, reject H_0 at the level of significance α if T exceeds $W_{1-\alpha}$,

$$W_{1-\alpha} = [n(n+1)/4] + X_{1-\alpha}\{[n(n+1)][(2n+1)/24]\}^{1/2} \qquad (2)$$

where $X_{1-\alpha}$ is the pth quantile of a standard normal random variable.

METHOD

The procedure for conducting a Linear Regression-Wilcoxon Test are given in the following steps. [For a discussion of the Wilcoxon Test, see Conover (2).]

1. Perform a linear regression on the time series data for the baseline period of the treatment area where the dependent variable is the evaluation metric of concern, and the independent variable is the numbered time period so that

$$y'_t = a_t + b_t X \qquad (3)$$

2. Perform a similar linear regression on the time series data for the baseline period of the control data so that

$$y'_c = a_c + b_c X \qquad (4)$$

3. Compute the standard error of estimate about the regression lines for both the control and treatment areas so that

$$S_t = \{(1/n-2) \sum_{i=1}^{M} [y_{t_i} - (a_t + b_t X_i)]^2\}^{1/2} \qquad (5)$$

$$S_c = \{(1/n-2) \sum_{i=1}^{M} [y_{c_i} - (a_c + b_c X_i)]^2\}^{1/2} \qquad (6)$$

4. Compute regression values for each time interval in the operational period for both the control and treatment areas so that

$$y'_{t_j} = a_t + b_t X_j \qquad \text{for } j = M + 1 \text{ to } N \qquad (7)$$

$$y'_{c_j} = a_c + b_c X_j \qquad \text{for } j = M + 1 \text{ to } N \qquad (8)$$

5. For each time period in the operational period, subtract the actual data point from the regression value for that period and area so that

$$d_{t_j} + y_{t_j} - y'_{t_j} \qquad \text{for } j = M + 1 \text{ to } N \qquad (9)$$

$$d_{c_j} = y_{c_j} - y'_{c_j} \qquad \text{for } j = M + 1 \text{ to } N \qquad (10)$$

6. Divide the values obtained in step 5 by the standard error of estimate for the respective area so that

$$Z_{t_j} = d_{t_j}/S_t \qquad \text{for } j = M + 1 \text{ to } N \qquad (11)$$

$$Z_{c_j} = d_{c_j}/S_c \qquad \text{for } j + M + 1 \text{ to } N \qquad (12)$$

7. Subtract the value obtained in step 6 for the treatment area from that obtained for the control area for each time period so that

$$D_j = Z_{c_j} - Z_{t_j} \qquad \text{for } j = M+1 \text{ to } N \qquad (13)$$

8. Rank the difference obtained in step 7 according to their absolute values.

9. Add the rankings of the positive differences. This is the test statistic (T).

10. If T exceeds

$$W_{1-\alpha} = [n(n+1)/4] + X_{1-\alpha}\,[n(n+1)(2n+1)/24]^{1/2}$$

where $X_{1-\alpha}$ is the $1-\alpha$th quantile of a standard normal random variable, and n = the number of time intervals in the operational period, then reject Ho at the α level of significance and conclude that data for the treatment area were significantly different from that of the control area.

EXAMPLE

Data are presented here to illustrate the evaluation technique. These data were taken from a pedestrian program for which the nature and the effectiveness have been reported elsewhere (3). The pedestrian program was of an educational nature involving 6- and 7-year-old children in four major cities of Alabama. Accident data for the 6- and 7-year olds constituted the treatment observations and accident data for the other age groups within the four-city area were used as the control. These data are presented in Tables 1 and 2, respectively. The pedestrian educational program began in the fall of 1978; therefore, October 1 was used as the program intervention point for the treatment area.

TABLE 1 Pedestrian Accident Data for 6-7 Year Age Group

Month	1976	1977	1978	1979	1980
January	2	2	6	6	5
February	3	7	6	0	1
March	6	3	6	0	3
April	5	5	7	3	5
May	11	7	5	3	3
June	3	8	3	4	6
July	4	3	2	1	6
August	5	4	8	3	3
September	6	5	7	5	4
			I		
October	7	4	8	2	-
November	6	6	3	2	-
December	3	3	2	4	-

Note: I = intervention. This refers to the point in time at which the pedestrian educational program began.

TABLE 2 Pedestrian Accident Data for All Other Age Groups

Month	1976	1977	1978	1979	1980
January	79	34	41	43	42
February	61	41	39	33	38
March	53	37	46	49	40
April	49	38	60	40	52
May	55	57	52	38	45
June	39	35	49	51	38
July	50	44	52	42	44
August	40	44	51	44	35
September	40	54	48	46	54
October	56	56	60	57	-
November	54	47	41	56	-
December	45	55	38	48	-

When the preceding technique was applied, the following regression equation was obtained for the treatment data:

$$y_t = 4.76894 + .02072\ X$$

with a standard error of estimate

$$S_t = 2.0963.$$

The regression equation for the control data was

$$y_c = 47.15150 + .02674\ X$$

with a standard error of estimate

$$S_c = 7.5727.$$

The remainder of the steps were carried out yielding a test value of

$$T = \Sigma R(+) = 227$$

which was statistically significant at the .05 level.

DISCUSSION

Few safety programs can be evaluated according to procedures used in a fully controlled laboratory experiment. In a field-type experiment, it is seldom possible to control, and often difficult to even identify and monitor, all the factors that could conceivably affect an experiment. Even when considerable time and effort are spent and the most sophisticated methods are used it still must be recognized that alternative explanations may exist and that the conclusions may not be as strong as if they had come from a laboratory experiment. On the other hand, a strong conclusion from a laboratory experiment may not be relevant in the real world where factors may be uncontrollable.

The technique presented herein is simple and easy to use in the evaluation of field experiments. However, as with the evaluation of any field experiment, considerable judgment must be used in the interpretation of significant results. Efforts must be made to identify and check as closely as possible all factors that could provide an alternative explanation to any effects found in the data.

ACKNOWLEDGMENT

Partial financial support for this research was provided by the Alabama Office of Highway and Traffic Safety.

REFERENCES

1. J.A. Caporaso and L.L. Roos. Quasi-Experimental Approaches Testing Theory and Evaluating Policy. Northwestern University Press, Evanston, Ill., 1973.
2. W.J. Conover. Practical Nonparametric Statistics. John Wiley and Sons, Inc., New York, 1971.
3. J.C. Fortenberry and D.B. Brown. Problem Identification, Implementation and Evaluation of a Pedestrian Safety Program. Accident Analysis and Prevention, Vol. 14, 1982, pp. 315-322.

Publication of this paper sponsored by Committee on Traffic Records and Accident Analysis.