FHWA/NJ-86-007-7788

CONCEPTS, APPLICATIONS, AND MISAPPLICATIONS OF THE CHI-SQUARE STATISTIC

Third Interim Report

Project 7788

Development of Statistical Procedures

Prepared By

Ricardo T. Barros

New Jersey Department of Transportation

Division of Research and Developement

In Cooperation With

U. S. Department of Transportation

Federal Highway Administration

November 20, 1985

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA/NJ-86/007		

4. Title and Subtitle	5. Report Date / 6. Performing Organization Code
Concepts, Applications, and Misapplications of the Chi-Square Statistic	5. November 20, 1985 6.
7. Author(s) Ricardo T. Barros	8. Performing Organization Report No. 86-007-7788
9. Performing Organization Name and Address New Jersey Department of Transportation Division of Research and Demonstration 1035 Parkway Avenue Trenton, NJ 08625	10. Work Unit No. 11. Contract or Grant No. NJ HPR Study 7788
12. Sponsoring Agency Name and Address Federal Highway Administration U.S. Department of Transportation Washington, DC 20590	13. Type of Report and Period Covered Interim Report #3 14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

Chi-square tests in a 2x2 matrix are, perhaps, too routinely applied. Their technical simplicity facilitates conceptual abuse by practitioners unfamiliar with the test's underlying constraints. This report reviews proper application of the chi-square test in a 2x2 table, cites a condition in which a 2x2x2 table may be more appropriate, and suggests an option available when cells in such tables are empty.

17. Key Words	18. Distribution Statement
Accident Statistics, Before and After Analyses, Chi-Square Statistic, Contingency Tables, Statistical Analysis of Accident Data, Structural Zero	No Restrictions

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	12	

Form DOT F 1700.7 (8-69)

CONTENTS

TABLES

FIGURES

CONCEPTS, APPLICATIONS, AND MISAPPLICATIONS OF THE CHI-SQUARE STATISTIC

INTRODUCTION

An engineer is interested to know whether a specific traffic maneuver is associated with accidents. The engineer realizes that site characteristics at various intersections differ and recognizes that some form of control will be necessary to screen these effects from this analysis. Historical accident data is gathered for two groups each of 50 intersections representing the one year period immediately preceding the experiment's start. The experimental maneuver is allowed at all test sites and these sites are posted accordingly. Signs prohibiting the maneuver are posted at all control sites. Accident counts are then recorded during the ensuing year and the data tabulated as shown below.

TABLE 1. A FOURFOLD CONTINGENCY TABLE

	ACCIDENT COUNT	
PERIOD	MANEUVER ALLOWED	MANEUVER PROHIBITED
BEFORE	16	20
AFTER	41	23

A quick reference to an elementary statistical text identifies this as a fourfold (2x2) contingency table and indicates that data thus tabulated is commonly evaluated using the chi-square distribution with one degree of freedom. The chi-square statistic is computed using Equation (1) and compared to the tabulated value associated with an appropriate level of risk, say $\chi^2_{1,\alpha=0.05}=3.841$.

$$X^2 = \sum(O-E)^2/E \tag{1}$$

It should be noted that, in practice, Equation (2) provides a more convenient variant to calculate the chi-square statistic in fourfold contingency tables.

$$X^2 = \frac{(a+b+c+d)(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} \tag{2}$$

The variables a, b, c, and d are simply the observed frequencies in each of the cells in the 2x2 matrix of Table 1. In designating these variables, it is important only that values a and d lie on the same diagonal.

Either of two practical outcomes may result from substituting the observed accident frequencies into Equation (2). If the X^2 statistic were to produce a value whose magnitude is greater than 3.841, then the engineer concludes that, with 95% confidence, accidents are in some way associated with the traffic maneuver and the maneuver is consequently prohibited at all intersections. Alternatively, if the X^2 statistic is nonsignificant, the engineer concludes that the traffic maneuver is safe and allows it universally.

Regardless of which action is taken, this engineer has joined the large group of unsuspecting researchers who have misapplied a statistical test. The logical errors, which may not be immediately apparent, will become more obvious if the possible tests which could have been construed appropriate are identified.

FOURFOLD CONTINGENCY TABLE ANALYSES

When data can be categorized in one of two ways in accordance with each of two criteria, a fourfold contingency table results. The three basic chi-square analyses appropriate to 2x2 contingency tables (1) are presented in Table 2 and examples of their use are presented in Table 3. Each of the presented analyses embodies a subtle distinction, however, predominantly reflecting the manner in which the data was collected. These distinctions may, in turn, narrow the selection of hypotheses which can be tested. It will be shown that the design assumptions prerequisite to the performance of any of these three analyses have not been satisfied by the hypothetical experiment previously described and, consequently, the engineer's statistical analysis supports neither of the conclusions previously suggested.

Of the three experimental designs presented in Table 2, it is immediately clear that assumptions B4 and C3 are not met by the manner in which the data was collected. Assumption C3 would require that a fixed number of accidents be submitted to this test and, of these, a predetermined number are to have occurred in each category for the Time and Maneuver classifications. In fact, the number of accidents in each row and column are the result of a random phenomenon and not fixed. Assumption B4 is similarly not met. B4 would have required that a fixed number of accidents be considered from both the Before and After periods when, in practice, the actual number of accidents therein were obviously unfixed variables.

The possible hypotheses to be tested are therefore limited by the manner in which the data was collected to that of analysis A. The unique hypothesis tested by analysis A is that of complete independence. Thus any finding of statistical significance here stops short of translating this inference to the specific effect

TABLE 2. DESIGN STRUCTURE OF 2x2 CHI-SQUARE CONTINGENCY TABLE ANALYSIS

FOURFOLD ANALYSIS	POSSIBLE NULL HYPOTHESES TO BE TESTED	REQUISITE ASSUMPTIONS
A	H_0: Independence	A1) The sample of N observations constitutes a random sample. A2) Each observation may be classified into exactly one of two categories according to a first criterion and exactly one of two categories according to a second criterion. A3) Both row and column totals are random.
B	H_0: Independence H_0: Equal probabilities.	B1) Each sample is a random sample. B2) The outcome of the various samples are mutually independent. B3) Each observation may be classified into exactly one of two column criteria. B4) Row totals are fixed and column totals are random.
C	H_0: Independence H_0: Equal probabilities	C1) Each observation is classified into exactly one cell. C2) Each observation is a random sample and has the same probability of being classified into cell (i,j) as any other observation. C3) Both row and column totals are fixed, not random.

Note: The hypothesis of independence is that the row classification is independent of the column classification. The hypothesis of equal probabilities is that all probabilities in the same column are equal to each other.

TABLE 3. 2X2 CONTINGENCY TABLE ANALYSES, ILLUSTRATIVE EXAMPLES

FOURFOLD ANALYSIS	ILLUSTRATIVE EXAMPLE	FOURFOLD TABLE			
			BLACK	WHITE	
A	Select 100 marbles at random and observe both their color and their condition.	CRACKED	a	b	y
		NOT CRACKED	c	d	(100-y)
			x	(100-x)	
			BLACK	WHITE	
B	Select 50 white marbles at random and another 50 black marbles at random. Observe their condition.	CRACKED	a	b	y
		NOT CRACKED	c	d	(100-y)
			50	50	100
			BLACK	WHITE	
C	Select 50 white marbles at random and another 50 black marbles at random, all in perfect condition. Subject all marbles to heat until 50 marbles crack.	CRACKED	a	b	50
		NOT CRACKED	c	d	50
			50	50	100

of an individual row or column. In other words, in the context of the examples in Table 3, a significant result in Analysis A may simply indicate that more black cracked marbles are to be found in the population at large, but not necessarily that black marbles crack more easily.

Furthermore, the experimental data collected by the engineer has been misrepresented by the contingency table he has constructed (Table 1). The experimental maneuver is never allowed under the Before condition. The upper left cell in the contingency table of Table 1 is a priori zero. This is referred to in the literature as a structural zero (2, p. 108). Analyses of contingency tables which contain structural zeros requires procedures slightly more complex than those commonly cited in introductory statistical texts and, in the case of a fourfold table, is not possible as the resultant degrees of freedom is zero.

At this point it should be emphasized that no stance has been taken as to whether the engineer was in fact correct or not in implementing the traffic restrictions. Rather, the point made is that the analysis performed to support his actions violates fundamental assumptions and potentially provides incorrect support for any subsequent conclusions. Specifically, (a) the experimental design used incorporates a structural zero which has not been recognized by the calculations performed, (b) the finding of statistical significance does not indicate that the experimental maneuver is either more or less dangerous, and (c) an unwarranted sense of security may have been conferred by the process of conducting a statistical test.

THE CHI-SQUARE STATISTIC AND HYPOTHESIS TESTS

Visual observation of the data in Table 1 suggests that accident frequencies are dramatically associated with the maneuver in question. The lower left quadrant exhibits what appears to be a disproportionately large increase in accident frequency. (This is, in fact, an accurate assessment as this simulated data was produced with a Poisson random generator in which the mean accident frequency more than doubled when the maneuver was allowed.) However, one can only reach such a conclusion if the appropriate statistical analysis is performed and the computed test statistic produces a significant result.

Assume, for the moment, that the problem of a structural zero may be ignored. The chi-square statistic is to be applied to the data in Table 1 in accordance with the most appropriate of the analyses listed in Table 2.

Note that both row and column totals must be considered random quantities -- they were not fixed prior to the experiment's start -- and that this fourfold table meets only the requirements for analysis A in Table 2. Thus, the only hypothesis which can be tested is that shown below.

H_o: Accident frequencies are independent of the Time and Maneuver category under which they may be classified.

H_o is to be accepted unless the computed X^2 statistic exceeds a predetermined critical value. Otherwise, one accepts the alternative hypothesis, H_A, as follows:

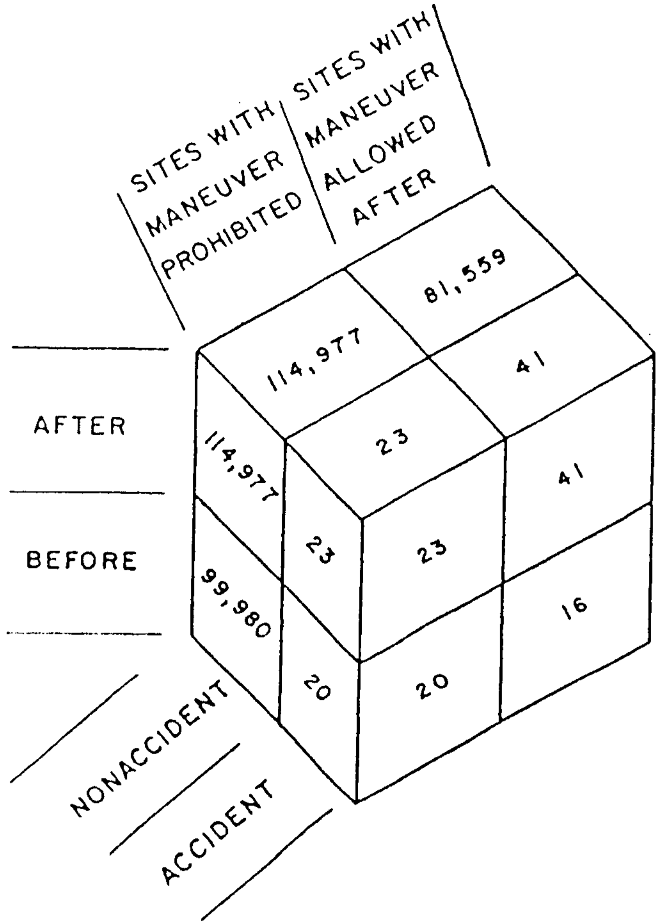
H_A: Accident frequencies are not independent of Time and Maneuver.

Perhaps surprisingly, at the 5 percent level of risk, the observed accident counts in this example are not sufficiently unusual to warrant the rejection of the null hypothesis and a substantial increase in accident potential has passed undetected. Computer simulation tests indicate it would pass undetected somewhat more than 50 percent of the time. Protection against this undesirable result requires that additional information be incorporated into the statistical test.

A brief description of the mathematical model used to generate the above accident data will quickly reveal the primary flaw in the preceding analysis. Observed accident counts were modeled with a Poisson random generator subject to the following constraints. Control sites sustained 100,000 exposures, i.e., 100,000 vehicles passed through these sites during the Before time period. Test sites sustained 80,000 exposures during the same period. Due to extraneous factors intentionally included in this particular simulation test, the exposure at the control sites increased by 15 percent by the end of the After time period, while the corresponding increase at the test sites was only 2 percent. Control and test sites were similar in all other respects. The probability of an accident was fixed at 1 in 5000 exposures if the maneuver was prohibited and 1 in 2000 exposures if the maneuver was allowed. This resulted in a model in which the accident frequencies in each quadrant of a 2 x 2 table had Poisson means of 23, 40.8, 20, and 16. While this 2 x 2 table may accurately reflect the constraints under which it was created, it is clearly insensitive to the varying exposure and its chi-square statistic cannot estimate accident probabilities, as demonstrated by the statistic's inability to reject the null hypothesis.

Addition of another layer to this table, reflecting all non-accidents which also occurred, will remedy this problem. The revised contingency table now has dimensions 2 x 2 x 2. This expanded matrix is shown in Figure 1a. Application of

(a) A complete table.



(b) A table with structural zeros.

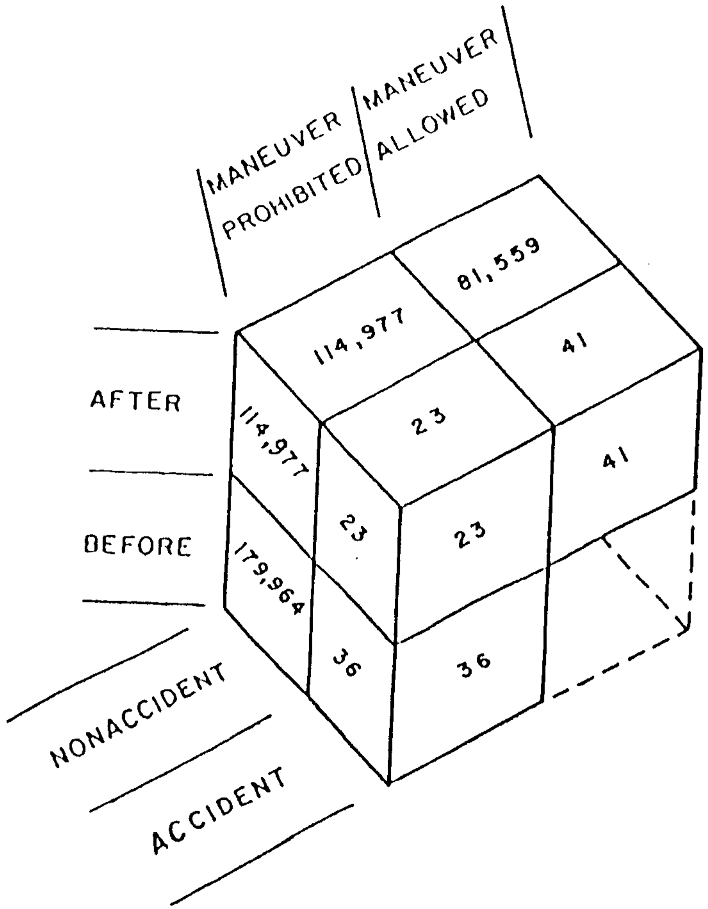


Figure 1. Examples of two 2x2x2 contingency tables.

the appropriate chi-square statistic -- a more sophisticated variant of Equation (1), but still ignoring the structural zero -- reverses the previous conclusion and correctly rejects the hypothesis of independence (3).

CONTINGENCY TABLES WITH STRUCTURAL ZEROS

Structural zeros are defined as those cells in a contingency table for which the expected and observed frequencies are a priori zero. In the context of the preceding example, the Before and the Allowed headings are indices of a structural zero. This is because the maneuver was never Allowed during the Before condition. Note that, as shown in Figure 1b, the accident data which previously appeared in these cells are now legitimately pooled into those cells with the Before-Prohibited indices.

The hypothesis of complete independence is tested with yet another variant of Equation (1) in which the expected values are obtained iteratively with computer assistance (3). The inferences made are, in this case, the same as those made with the more conventional 2x2x2 analysis, i.e., the accident frequencies are not independent of Time and Maneuver. However, simulation tests indicate this technique is slightly more powerful than the conventional procedure in rejecting the null hypothesis when it is truly false.

Whether the net effect of this consideration warrants the more complex calculations involved may depend largely on the availability of computer assistance. If it is readily available, then additional discriminating power may be obtained with minimal additional effort. Otherwise the more conventional 2 x 2 x 2 analysis may be employed to obtain very nearly the same results. In either case, one thing is clear -- either of these methods is preferable to that of the fourfold table with unconstrained marginal totals. Application of the routine fourfold contingency table analysis to the above data correctly rejected the null hypothesis only 47 percent of the time in a computer simulation check.

SUMMARY AND CONCLUSIONS

A statement whose impact may now be more fully appreciated is the following:

"It is unfortunate that the chi-square statistic takes such a simple form, both because its calculation does not require the investigator to determine explicitly the proportions being contrasted and because it invites the investigator to ignore the fact that the proper inferences to be drawn from the magnitude of X^2 depends on how the data was generated, even though the formula for X^2 does not" (4).

Both pitfalls cited in this quotation were encountered by the hypothetical engineer described earlier in this paper. Using only raw data, exposure was unwittingly overlooked and, also, the inferences drawn were based on a hypothesis not under test. Furthermore, the existence of a structural zero in the table investigated passed unacknowledged.

In a very limited sense, the X^2 statistic is forgiving. Relatively large effects are required in a fourfold table before a significant statistic can be consistently produced. This results in a situation where the null hypothesis is accepted too frequently whenever the differences sought are relatively small. Consequently, the alternative hypothesis is accepted less frequently and the errors that might have been made have a reduced opportunity within which to manifest themselves. But this must not be construed to compensate for logical errors. Rather, it simply makes them less visible.

REFERENCES

1. Conover, W. J., Practical Nonparametric Statistics, 2nd Edition, John Wiley and Sons, New York, 1980.

2. Everitt, B. S., The Analysis of Contingency Tables, John Wiley and Sons, New York, 1977.

3. Barros, R. T., Before-After Analyses and the Chi-Square Statistic, Draft Report. New Jersey Department of Transportation, 1985.

4. Fleiss, J. L., Statistical Methods for Rates and Proportions, John Wiley and Sons, New York, 1973, p. 15.