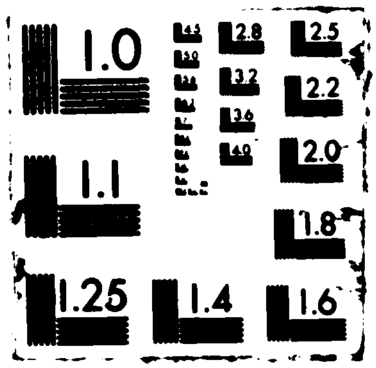
AD-A185 004 DETERMINING DIFFERENCES IN RATES CORRESPONDING TO A GIVEN SIGNIFICANCE LEVEL(U) NAVAL RESEARCH LAB WASHINGTON DC A R MILLER ET AL. 04 SEP 87 NRL-MR-6060 1/1

UNCLASSIFIED F/G 12/3 NL


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0

aval Research Laboratory
hington, DC 20375-5000

NRL Memorandum Report 6060

Determining Differences in Rates Corresponding to a Given Significance Level

ALLEN R. MILLER

Engineering Services Division

MARTIN FEUERMAN

New Jersey Medical School
Newark, N.J. 07103

September 4, 1987

ADA185004

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY 2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release, distribution is unlimited.
4. PERFORMING ORGANIZATION REPORT NUMBER(S) NRL Memorandum Report 6060	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Naval Research Laboratory 6b. OFFICE SYMBOL (If applicable) 2303	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) Washington, D.C. 20375-5000	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING / SPONSORING ORGANIZATION 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)

Determining Differences in Rates Corresponding to a Given Significance Level

12. PERSONAL AUTHOR(S)

Miller, Allen R. and Feuerman, Martin

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM Oct 86 TO Mar 87	1987 September 4	10

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Chi-square test Statistical significance Statistics

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The well-known Chi-square test can be used to determine whether the difference in two rates is statistically significant, and if so, at what level of significance. This paper discusses the related question of how large a difference in rates must be (when one rate is held constant) in order to show statistical significance at a given level of significance. An illustrative example, adapted from data appearing in the biostatistical literature, is provided.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED	
22a. NAME OF RESPONSIBLE INDIVIDUAL Allen R. Miller	22b. TELEPHONE (Include Area Code) (202) 767-2215	22c. OFFICE SYMBOL 2303

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted. All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE

Determining Differences in Rates Corresponding to a Given Significance Level

Introduction

The well-known Chi-square test can be used to detect significant differences between two proportions. For example, in Table I, we have N patients, broken down by (A + B) who received a placebo, and (C + D) patients who received treatment. The proportion of control patients who recovered is $A/(A+B)$, and the proportion of treated patients who recovered is $C/(C+D)$.

	Recovery	No Recovery	Totals
Controls	A	B	A + B
Treated	C	D	C + D
Totals	A + C	B + D	N = A + B + C + D

Table I

To determine if there is a statistically significant difference between these proportions, one uses the Chi-square test (formula) for 2×2 tables. For moderate to large N, many statistical computer packages (e.g. SPSS-X) utilize the formula with Yates correction, that is,

$$\chi^2 = \frac{N(|AD - BC| - N/2)^2}{(A+B)(C+D)(A+C)(B+D)} \tag{1}$$

Manuscript approved March 11, 1987.

Let us assume that we have performed a Chi-square test, and we do <u>not</u> get a significant result at say, the 5% level. In otherwords, suppose the proportion who recovered among the treated patients is larger than the proportion who recovered among the control patients (i.e., $C/(C+D) > A/(A+B)$), but the difference in these proportions is not significant at the 5% level (i.e., $\chi^2 < 3.84$). Now if a result is significant at the 5% level, we are implying that the probability is less than 5% that the result is due to chance. One might ask the following question: Assuming constant A, B, and N, how large should the recovery rate of the treated be in order to achieve significance at the 5% level (i.e., $\chi^2 \geq 3.84$)?

Modifying Table I, we arrive at Table II.

	Recovery	No Recovery	Totals
Controls	A	B	A + B
Treated	y	N - (A+B+y)	N - (A + B)
Totals	A + y	N - (A+y)	N

Table II

We now discuss how to find the treated recovery rate $y/[N - (A + B)]$.

Computing the Recovery Rate

Applying the Chi-square formula (1) to Table II, we obtain

$$\chi^2 = \frac{N\{|A[N-(A+B+y)] - By| - N/2\}^2}{(A+B)[N-(A+B)](A+y)[N-(A+y)]} .$$

If we assume that the recovery rate is higher among the treated, that is,

$$\frac{y}{N-(A+B)} > \frac{A}{A+B} ,$$

then the quantity inside the absolute value signs is negative. Consequently χ^2 can be rewritten as

$$\chi^2 = \frac{N\{(A+B)y - A[N-(A+B)] - N/2\}^2}{(A+B)[N-(A+B)](A+y)[N-(A+y)]} ,$$

which leads to a quadratic equation in y.

Defining

$\xi \equiv A + B$,

$\eta \equiv N - \xi$,

$\mu \equiv N - A$, and

$a \equiv A\eta + N/2$,

we obt. n

$$\alpha y^2 - \beta y + \gamma = 0 \tag{2}$$

where

$$\alpha \equiv N\xi + n\chi^2 \ ,$$

$$\beta \equiv 2Na + n\chi^2(\mu - A) \ , \text{ and}$$

$$\gamma \equiv Na^2/\xi - n\chi^2 A\mu \quad .$$

Let us assume for the moment that Eq. (2) has real roots. This will be verified later by showing that the discriminant of the quadratic is positive.

Now, even for very small levels of significance (p-values) of say, .001 and .0001, χ^2 will assume values only as large as 10.8 and 15.1, respectively. So we can safely assume that $N > \chi^2$. (The Chi-square test with Yates' correction for 2×2 tables is generally not used for tables where N is 20 or less.) This assumption implies that $\beta > 0$. Clearly, $\alpha > 0$, so that the sum of the roots, $\beta/\alpha > 0$.

In addition, if we define

$$\delta \equiv \frac{A+1}{A+\chi^2} N - \xi \quad , \tag{3}$$

it is possible to show that the condition $\delta \geq 0$ implies that $\gamma > 0$. Therefore, under this condition, the product of the roots $\gamma/\alpha > 0$, and we are insured of two positive roots. Regardless of the value of δ, we can be certain that Eq. (2) will have at least one positive root.

We now show that the discriminant of the quadratic is positive, so we can be certain that the roots are real. The discriminant, $\Delta \equiv \beta^2 - 4\alpha\gamma$ can be written after a straightforward computation:

$$\Delta = (\eta N\chi^2)^2 + 4\eta N^2\chi^2[A + \frac{a}{\xi}][B - \frac{1}{2}] \quad .$$

We observe $B \geq 1$. Otherwise, if $B = 0$ we would have 100% of the controls recovering. Consequently $\Delta > 0$ and the roots of Eq. (2) are real. As shown earlier, we can also conclude that at least one of the roots will be positive. Before giving an example, we note finally that any positive roots that are found must satisfy the inequality

$$\frac{A}{A+B} < \frac{y}{N-(A+B)} \leq 1 \quad , \tag{4}$$

i.e., the recovery rate of the treated must be greater than the rate for the controls, and not more than 100%.

Illustrative Example

Table III is adapted from data that appears in Chinn [1], which is also discussed in Bliss [2].

	Well	Sea-Sick	Totals
Placebo	18	12	30
Dramamine	y	34 - y	34
Totals	18 + y	46 - y	64

Table III

In this example, $A = 18$, $B = 12$, $N = 64$, $\xi = 30$, $\eta = 34$, $\mu = 46$, $\mu - A = 28$, $a = 644$, $Na^2/\xi = 884770.13$, and $\chi^2 = 3.84$ (i.e., corresponding to the 5% level of significance). Equation (2) then becomes

$$2050.56y^2 - 86087.68y + 776666.45 = 0$$

Using Eq. (3) we find that $\delta = 25.68$ so we expect two positive roots. Solving the quadratic, we obtain $y = 28.86$ and 13.12. Since inequality (4) must be satisfied, we reject the root $y = 13.12$, and accept the root $y = 28.86$. In actuality, our solution must be in integers, and we obtain $[y] + 1 = 29$. Thus, in order to show a statistically significant difference in the rates at the 5% level, at least 29 of the 34 treated patients must recover (i.e., stay well).

Conclusions

The problem discussed in this paper arises when one does not get a significant result when comparing the difference in two proportions (rates) in a 2×2 contingency table. The research worker might be interested in determining how large a difference in rates is required (assuming one rate is held fixed) in order to show statistical significance at a given level. This paper provides a method for computing the required difference for an arbitrary level of statistical significance. An illustrative example applying the method is given, based on data appearing in the literature.

References

[1] Chinn, H. I., W. K. Noel, and P. K. Smith, "Prophylaxis of motion sickness: evaluation of some drugs in seasickness", USAF School of Aviation Medicine, Project 21-32-014, Rep. 4, 1950.

[2] Bliss, C. I., Statistics in Biology, Volume I, New York, McGraw-Hill, p. 65, 1967.

END

11-87

DTIC