NONASYMPTOTIC INFERENCES BASED ON
COCHRAN'S Q TEST

PAUL W. MIELKE, JR. AND KENNETH J. BERRY
Colorado State University

---

1 Request reprints from P. W. Mielke, Jr., Department of Statistics, Colorado State University, Fort Collins, CO 80523.
NONASYMPTOTIC INFERENCES BASED ON COCHRAN'S Q TEST

Summary.—A nonasymptotic inference procedure for Cochran's Q test for the equality of matched proportions is presented. The nonasymptotic method provides improvement over the asymptotic method when there is a small number of subjects and/or a large number of treatments.
NONASYMPTOTIC INFERENCE BASED ON COCHRAN'S

Q TEST

Cochran's (1950) Q test for the equality of matched proportions is widely used in educational and psychological research. The test can be viewed as an extension of the McNemar (1947) test to three or more treatment conditions. As an example application, suppose that a sample of \( n \geq 2 \) subjects are observed in a situation where each subject performs individually under each of \( g \geq 2 \) different experimental conditions. Each subject is assigned the treatment conditions in a random order and, within each condition, each of the \( n \) subjects performs one of a set of \( g \) motor skill tasks. If the task is completed in a fixed time period, the performance is scored as a success (1) and as a failure (0) otherwise. The research interest of the experimenter is to evaluate if the motor skill tasks are of equal difficulty for the subjects, i.e., is the true proportion of successes constant over the \( g \) conditions?

A nonasymptotic approach to this problem which is based on Cochran's Q test is presented in this paper. The nonasymptotic method provides improved analyses for those cases when the asymptotic approach may be questionable, e.g., a small number of subjects, a large number of treatment conditions, and/or a relatively small proportion of successes for subjects.
Cochran's $Q$ test statistic for the analysis of $g$ treatment conditions (columns) and $n$ subjects (rows) is given by

$$Q = \frac{(g-1)(g \sum_{j=1}^{g} C_j^2 - A_1^2)}{gA_1 - A_2} \tag{1}$$

where

$$A_m = \sum_{i=1}^{n} R_i^m, \tag{2}$$

$$R_i = \sum_{j=1}^{g} X_{ij} \tag{3}$$

is the number of 1's in the $i$th of $n$ rows,

$$C_j = \sum_{i=1}^{n} X_{ij} \tag{4}$$

is the number of 1's in the $j$th of $g$ columns, and $X_{ij}$ denotes the cell entry of either 0 or 1 associated with the $i$th of $n$ rows and the $j$th of $g$ columns. The null hypothesis ($H_0$) underlying the nonasymptotic approach stipulates that each of the
\[ M = \prod_{i=1}^{n} \left( \frac{g}{R_i} \right) \]  

[5]

distinguishable configurations of 1s and 0s within each of the \( n \) rows, given that the values \( R_1, \ldots, R_n \) are fixed, occurs with equal probability.

The asymptotic approach specifies that the distribution of \( Q \) under \( H_0 \) is approximated by the chi-squared distribution with \( g-1 \) degrees of freedom. The well-known asymptotic mean, variance, and skewness of \( Q \) are given by

\[ \mu_Q = g-1, \]  

[6]

\[ \sigma_Q^2 = 2(g-1), \]  

[7]

and

\[ \gamma_Q = \left( \frac{8}{g-1} \right)^{1/2}, \]  

[8]

respectively.

In contrast, the nonasymptotic approach is based on the conditional permutation distribution of the \( M \) distinct tables under \( H_0 \). Thus, the exact mean, variance, and skewness of \( Q \) under \( H_0 \) are given by
\[ \mu_0 = g - 1, \quad [9] \]

\[ \sigma_0^2 = \frac{2(g-1)B_1}{g(A_1-A_2)^2}, \quad [10] \]

and

\[ \gamma_0 = \frac{\kappa_3(Q)}{\sigma_0^3} \quad [11] \]

where

\[ \kappa_3(Q) = \frac{4(g-1)}{(gA_1-A_2)^3} \left[ \frac{g-1}{g-2} B_2 \theta(g) + 2B_3 \right], \quad [12] \]

\[ B_1 = g^2(A_1^2 - A_2) - 2g(A_1A_2 - A_3) + (A_2^2 - A_4), \quad [13] \]

\[ B_2 = g^4(A_1^2 - A_2) - 6g^3(A_1A_2 - A_3) + g^2(4A_1A_3 + 9A_2^2 - 13A_4) \]
\[ - 12g(A_2A_3 - A_5) + 4(A_3^2 - A_6), \quad [14] \]

\[ B_3 = g^3(A_1^3 - 3A_1A_2 + 2A_3) - 3g^2(A_1^2A_2 - 2A_1A_3 - A_2^2 + 2A_4) \]
\[ + 3g(A_1A_2^2 - A_1A_4 - 2A_2A_3 + 2A_5) - (A_2^3 - 3A_2A_4 + 2A_6), \quad [15] \]
and \( \Theta(g) = 0 \) if \( g = 2 \) and \( \Theta(g) = 1 \) if \( g \geq 3 \).

Given the exact values of \( \mu_\varphi \), \( \sigma_\varphi^2 \), and \( \gamma_\varphi \) under \( H_0 \), the nonasymptotic inference is based on the standardized statistic given by

\[
Z = \frac{Q - \mu_\varphi}{\sigma_\varphi} \quad [16]
\]

where the exact permutation distribution of \( Z \) is approximated by the Pearson type III distribution. The density function of the Pearson type III distribution is given by

\[
f(w) = \frac{(2/\gamma_\varphi)^{2/\gamma_\varphi^2}}{\Gamma(4/\gamma_\varphi^2)} \left[ (2 + \gamma_\varphi w)/\gamma_\varphi \right]^{(4 - \gamma_\varphi^2)/\gamma_\varphi^2} e^{-2(2 + \gamma_\varphi w)/\gamma_\varphi^2} \quad [17]
\]

where \(-2/\gamma_\varphi < w < \infty \) (Harter, 1969).

Consider an example involving \( n = 10 \) elementary school students who are tested on a series of \( g = 5 \) motor skill tasks (e.g., beam walking, ball throwing, etc.). Table 1 contains the recorded successes (1) and failures (0) together with the subject totals (\( R_i \) values) and task totals (\( C_j \) values). The results of the nonasymptotic analysis are \( Q = 9.3793 \), \( \mu_\varphi = 4.0000 \), \( \sigma_\varphi^2 = 7.1914 \), \( \gamma_\varphi = 1.1928 \), and the probability under \( H_0 \) of a \( Q \) value \( \geq 9.3793 \) is 0.0443.
REFERENCES


### TABLE 1

**SUCCESSES (1) AND FAILURES (0) OF n = 10 ELEMENTARY SCHOOL CHILDREN ON A SERIES OF g = 5 MOTOR SKILL TASKS**

<table>
<thead>
<tr>
<th>Subject</th>
<th>Motor Skill Task</th>
<th>( R_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

\( C_j \) 4 7 7 3 1