STATISTICAL ANALYSIS OF BINARY DATA FROM MATCHED DESIGNS
THAT REQUIRE MULTIPLE OBSERVATIONS ON THE SAME HERD, LITTER OR ANIMAL

ALAN DONALD

To determine the effectiveness of a vaccine for bovine respiratory disease, an investigator randomizes feedlot cattle to vaccine and control groups and then uses a chi-square test to compare morbidity rates. To control for confounding factors such as the environment, he may assign one half of each pen to a treatment. But here some care is needed. The chi-square test is valid only if the individual animal responses are independent: the similarity of responses within each pen may bias the test statistic and yield invalid results. This paper discusses this problem, which is called clustering, and reports on a computer simulation of three strategies for dealing with it in matched designs.

Clusters in veterinary research may be herds, pens, litters or even repeated observations on the same animal. The phenomenon of clustering may occur whenever observations are likely to be correlated.

We may model clustering in binary data by assuming that each cluster has its own probability of an event and defining the intrACLuster correlation coefficient, \( p \), as the proportion of the variance in the population that is due to variation between clusters. If there is a wide variation, the clustering is high and \( p = 1 \). Low variation (\( p = 0 \)) indicates no clustering. A detailed description of the phenomenon appears in Donald and Donner (1987).

MATCHED CLUSTERS

If clusters such as litters or pens are split between two treatments, the animals within each half cluster will exhibit similar responses. There will also be a similarity, modified by the treatment effect, between members of the same cluster on different treatments. This situation may be described in terms of the clustering model. Suppose a new vaccine is being tested and the mean rates of success (non-disease) are \( P_1 \) in the control group and \( P_2 \) in the treatment group. The odds ratio of the treatment effect is \( \psi = P_2 Q_1 / P_1 Q_2 \), where \( Q_1 = 1 - P_1 \) and \( Q_2 = 1 - P_2 \). Under the model, the control half of cluster \( i \) has its own success probability, \( P_{1i} \). The relation between \( P_{1i} \) and the probability, \( P_{2i} \), of success in the treated half of the cluster is described by:

\[
\psi = \frac{P_{2i} q_{1i}}{P_{1i} q_{2i}}.
\]

So \( P_{2i} \) can be expressed as

\[
P_{2i} = \frac{\psi P_{1i}}{\psi P_{1i} + q_{1i}}.
\]

1Department of Health Management, Atlantic Veterinary College, University of Prince Edward Island, Charlottetown, Prince Edward Island, Canada
By reducing the treatment effect to the cluster level, Eq. (1) captures the effect of both the treatment and the intracluster correlation.

METHODS OF ANALYSIS

Suppose the ith of n clusters (litters or herds) has \( m_i \) members and is split into a control and treatment groups containing \( m_{1i} \) and \( m_{2i} \) members respectively. Denote the number of successes (healthy animals) in group j by \( x_{ji} \) and the number of failures (diseased animals) by \( y_{ji} \). Let \( p_{ij} = x_{ji} / m_{ij} \) be the observed proportion of successes in cluster i group j and let \( p_i = (x_{1i} + x_{2i}) / m_i \) be the pooled proportion of successes.

We consider three methods of analysing data from split cluster designs.

The Mantel-Haenszel test

The Mantel-Haenszel chi-square test (Mantel and Haenszel, 1959) was designed to deal with binary data stratified by age, sex and breed in which the probabilities of success within each stratum are fixed. In this paper, each cluster (pair of half-clusters) forms a stratum and the success probabilities are random. The Mantel-Haenszel test, however, is widely available and has been shown to be robust under a variety of circumstances. It is performed by comparing

\[
X_{MH}^2 = \left( \frac{\sum_{i} \left( x_{1i}y_{2i} - x_{2i}y_{1i} \right)}{m_{ij}} \right)^2
\]

with the chi-square distribution with one degree of freedom.

The paired t-test

A t-test of the means of the observed differences, \( d_i = \hat{p}_{1i} - \hat{p}_{2i} \), may be used to test the null hypothesis of no treatment effect. Let \( d = \sum d_i / n \) and \( S^2 = \sum (d_i - \bar{d})^2 / (n - 1) \). The t statistic \( t_p = d / \sqrt{S} \) is referred to the t-distribution with \( n - 1 \) degrees of freedom.

The logistic transformation t-test

A logistic transformation of the observed differences in proportions is given by \( l_i = \ln((x_{1i} + .5)(y_{2i} + .5)/(x_{2i} + .5)(y_{1i} + .5)) \). This can be weighted by the reciprocal of its variance.

\[
u_i = \left( \frac{(m_{1i}+1)(m_{1i}+2)}{m_{1i}(x_{1i}+1)(y_{1i}+1)} \right) \left( \frac{(m_{2i}+1)(m_{2i}+2)}{m_{2i}(x_{2i}+1)(y_{2i}+1)} \right)^{-1}
\]

Let \( T = \Sigma u_i / \Sigma u_i \) and \( S^2 = \Sigma u_i(l_i - T)^2 / \Sigma u_i \). The test statistic, referred to the t-tables with \( n - 1 \) degrees of freedom, is

\[
t_l = \frac{T \Sigma u_i}{\sqrt{S_1 \Sigma u^2_i}}
\]
SIMULATION STUDY

To compare the qualities of the three tests, a simulation was carried out for the combinations of parameters (cluster size, m; odds ratio, \( \psi \); intracluster correlation, \( \rho \)) shown in Table 1. Each mean rejection rate represents 500 simulated experiments at each of the following values of \( P_1: .1, .2, .3, .4, \) and \( .5 \). Each experiment involved 10 clusters of size \( 2m \). For completeness the Pearson chi-square was included among the tests. Ideally, the rejection rates when \( \psi = 1 \) should be .05. High rejection rates when \( \psi = 2 \) indicate high power.

Table 1. Observed mean rejection rates of the four tests

<table>
<thead>
<tr>
<th>( \rho = )</th>
<th>.0</th>
<th>.1</th>
<th>.2</th>
<th>.3</th>
<th>.4</th>
<th>.5</th>
<th>( m = 5 )</th>
<th>( m = 25 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-square</td>
<td>.048</td>
<td>.045</td>
<td>.038</td>
<td>.029</td>
<td>.018</td>
<td>.011</td>
<td>.050</td>
<td>.046</td>
</tr>
<tr>
<td>Mantel-Haenszel</td>
<td>.020</td>
<td>.028</td>
<td>.029</td>
<td>.022</td>
<td>.016</td>
<td>.012</td>
<td>.040</td>
<td>.048</td>
</tr>
<tr>
<td>Paired t-test</td>
<td>.044</td>
<td>.046</td>
<td>.036</td>
<td>.036</td>
<td>.028</td>
<td>.022</td>
<td>.061</td>
<td>.051</td>
</tr>
<tr>
<td>Logistic t</td>
<td>.048</td>
<td>.053</td>
<td>.056</td>
<td>.063</td>
<td>.062</td>
<td>.058</td>
<td>.062</td>
<td>.054</td>
</tr>
<tr>
<td>Chi-square</td>
<td>.356</td>
<td>.286</td>
<td>.219</td>
<td>.150</td>
<td>.118</td>
<td>.072</td>
<td>.914</td>
<td>.853</td>
</tr>
<tr>
<td>Mantel-Haenszel</td>
<td>.276</td>
<td>.238</td>
<td>.196</td>
<td>.157</td>
<td>.129</td>
<td>.103</td>
<td>.905</td>
<td>.854</td>
</tr>
</tbody>
</table>

In small clusters (litters) the logistic t was the only test to resist the influence of high clustering when \( \psi = 1 \). In large clusters, the Mantel-Haenszel and paired t-test were slightly conservative under the null hypothesis for \( \rho \geq 0.1 \); the logistic t was anti-conservative. The power of the logistic t is slightly higher in small clusters when \( \psi = 2 \). For larger clusters the Mantel-Haenszel is slightly better than the logistic t. The low rejection rates in the small cluster simulation are due to the small sample size.

The results of the simulation study indicate that for studies involving split litters, the logistic t test should be used. For larger clusters (herds), the Mantel-Haenszel or logistic t perform well. The ordinary chi-square statistic is not recommended if clustering is suspected.

A sheet of worked examples is available from the author.

REFERENCES

Donald, Alan and Donner, Allan (1987). Adjustments to the Mantel-Haenszel chi-square statistic and odds ratio variance estimator when the data are clustered. Statistics in Medicine 6, 491-499.


448

Proceedings of the 5th International Symposium on Veterinary Epidemiology and Economics, 1988
Published in Acta Veterinaria Scandinavia, Supplementum 84
Available at www.sciquest.org.nz