Journal of Education and Practice
ISSN 2222-1735 (Paper) ISSN 2222-288X (Online)
Vol.4, No.23, 2013

Cochran–Mantel–Haenszel Test for Repeated Tests of Independence: An Application in Examining Students' Performance

Adejumo, A. O.
Department of Statistics, University of Ilorin, Kwara State, Nigeria
aodejumo@unilorin,edu.ng
Adetunji, A. A.
Department of Mathematics and Statistics, Federal Polytechnic, Ado-Ekiti, Ekiti State, Nigeria
adecap4u@yahoo.com

Abstract
From the result of graduate of ten departments in Faculty of Science, University of Ilorin for 2011/2012 academic session, data on final cumulative grade point average (Final Grade); department (ten departments of the faculty); age at entry (below or 20 years and above 20 years) and sex (male and female) are analyzed using Cochran-Mantel-Haenszel statistics. Odds of a student graduating with Second Class Upper and above (0.5270) is about half of graduating with Second Class Lower and below. This implies that the final grade is approximately symmetrical about two groups. The first group are those with Second Class Lower and below (Low Grade) while the other is for those with Second Class Upper and above (High Grade). Breslow-Day and Tarone's statistics show that the null hypothesis of homogeneity of odds ratio across the departments is not rejected for both age at entry and sex. This implies that the odds ratio across the ten departments (relating to age at entry & final grade and sex & final grade) are all equal. Cochran's and Mantel-Haenszel statistics reveals the final grade of students (Low Grade or High Grade) is not associated with both sex and age students at entry. The odds in favour of a student whose age is less than 20 years graduating with Low Grade (Pass, Third Class, and Second Class Lower) is 0.865 while it is 0.670 for male students graduating with lower grade.
Keywords: Test of Independence, Students' Performance, Cochran-Mantel-Haenszel

Introduction

Often, associations between two categorical variables are examined across two or more populations. The resulting data usually lead to several (say, H) 2 x 2 contingency tables. In many cases, the primary question involves the relationship between an independent variable (factor) that is either present or absent and a dependent (response) variable that is either present or absent in the presence of several covariates. This could give rise to frequency data that may be summarized as a set of 2 x 2 tables. In general, we are interested in collecting information for each of several 2 x 2 tables across the levels of the subpopulations (which may be determined by various configurations of factor variables or covariates).

Examining the cumulative odds of graduates of University of Ilorin, 2011/2012 academic session, the final grade of students are categorised in to *High Grade* (First Class and Second Class Upper) and *Low Grade* (Pass, Third Class, and Second Class Lower), see table 1. This table comes from a single study that has been stratified by a factor (ten departments in the faculty of science). The goal is usually to be able to combine the tables in order to have unified information across the tables. We would like to combine the evidence from the ten departments to make an overall statement about whether final grade is independent of sex (or independent of age at entry). The conditional test for these data within each department can be obtained by computing Fisher's exact test separately for each sub-table or obtain Pearson's X^2 for each table. Thus, the value of X^2 is first computed for each department, and the results suggest that neither of them is significant, which suggests that final grade is independent of sex (and age at entry) within each department.

The result of Pearson's X^2 and Fisher's exact test suggest that it would not be wise to collapse the data over the factor variable (department) without serious distortion to the association between the two variables been considered (Lawal, 2003).

Methodology

Design Considerations for a Contingency Table Analysis

Two Sampling Strategies

Two separate sampling strategies lead to the chi-square contingency table analysis.

1. *Test of Independence.* A single random sample of observations is selected from the population of interest, and the data are categorized on the basis of the two variables of interest. Such a sampling strategy would indicate that a single random sample of subjects of interest are selected, and each selected subject is categorized according to

factor of interest like sex, class of degree, mode of entry, etc.
2. *Test for Homogeneity.* Separate random samples are taken from each of two or more populations to determine whether the responses related to a single categorical variable are consistent across populations. For example, the interest may be to see the performance of students across various departments, random selection of subjects (students) from each of the departments may lead the researcher to determine whether the classes of degree differ among various departments.
The two-way table is set up the same way regardless of the sampling strategy, and the chi-square test is conducted in exactly the same way. The only real difference in the analysis is in the statement of the hypotheses and conclusions.
In Cochran's original 1954 paper, a test statistic was introduced to extend the chi-square test of independence in a 2 X 2 table to multiple 2 X 2 tables where each table corresponds to a different level of an intervening variable. Cochran proposed a test of conditional independence-independence of the variables forming the rows and columns of the tables, conditional on the levels of a third variable.
To establish notation, let n_{hij} represent the number of responses observed at the i^{th} level of the row variable, the j^{th} level of the column variable, and the h^{th} level of the intervening variable. Assuming H levels of the intervening variable $I = 2$ rows, and $J = 2$ columns, we have data that may be summarized as in the table a below for $h = 1, \ldots, H$.

	Factor B		**Total**
Factor A	n_{h11}	n_{h12}	$n_{h1.}$
	n_{h21}	n_{h22}	$n_{h2.}$
Total	$n_{h.1}$	$n_{h.2}$	$n_{h..}$

For this situation Cochran conditioned on the row totals, considering each 2 X 2 table to consist of independent binomials. He based his statistic on a weighted sum of the table-specific differences in proportions:

$$d_W = \sum_{h=1}^{H} [w_h(\hat{p}_{h1}] - \hat{p}_{h2}) \tag{1}$$

where $w_h = \frac{n_{h1.} * n_{h2.}}{n_{h..}}$ and $\hat{p}_{hi} = \frac{n_{hi1}}{n_{hi.}}, i = 1, 2$.
Using the asymptotic normality of d_w, he justified

$$\chi_C^2 = \frac{d_W^2}{\widehat{var}(d_W)} = \frac{d_W^2}{\sum_{h=1}^{H} \frac{(n_{h1.} * n_{h2.} * n_{h.1} * n_{h.2})}{(n_{h..}^3)}} \tag{2}$$

as an appropriate test statistic, having an approximate $\chi^2(1)$ distribution under the null hypothesis of conditional independence.
Mantel and Haenszel (1959) proposed a similar test statistic using a hypergeometric assumption. Conditional on the row and column totals, the cell counts in each table have a hypergeometric distribution. This fact suggests a test statistic based on the difference between the observed and expected frequencies in each 2 X 2 table. As with the classic chi-square test of independence in a single 2 X 2 table, it suffices to compare the observed and expected count in one cell per table (Bickel and O'Connell, 1975).
In particular, let n_{h11} be the "pivot" cell frequency of subjects in the *i-th* table who have both factor and response present. Under the assumption that the marginal totals are fixed, the overall null hypothesis of no partial association against the alternative hypothesis that on the average across the h sub-tables, there is a consistent relationship between the row and column variables is conducted by obtaining the Cochran-Mantel-Haenszel (CMH) test statistic χ_{MH}^2 , which is computed as follows:
For table *i,* for instance, $n_{h..} = (n_{h11}, n_{h12}, n_{h21}, n_{h22})$ follows the hypergeometric distribution and therefore

$$P\{n_{h..}|H_0\} = \left(\frac{n_{h1.}! * n_{h2.}! * n_{h.1}! * n_{h.2}!}{n_{h..}! * n_{h11}! * n_{h12}! * n_{h21}! * n_{h22}!}\right) \tag{3}$$

Hence, it follows that the expected value for the pivot cell in the *i-th* sub-table is given by:

$$E(n_{h11}|H_0) = \left(\frac{n_{h1.} n_{h.1}}{n_{h..}}\right) \tag{4}$$

$$Var(n_{h11}|H_0) = \left(\frac{n_{h1.}n_{h.1} * n_{h2.}n_{h.2}}{n_{h..}^2(n_{h..} - 1)}\right) \quad (5)$$

The Mantel-Haenszel test is, therefore,

$$\chi^2_{MH} = \frac{\left(\left|\sum_h \frac{n_{h1}n_{h2}}{n_{h.}}(p_{h1} - p_{h2})\right| - 0.5\right)^2}{\sum_h \frac{n_{h1}n_{h2}}{n_{h.} - 1}\overline{pq}} \quad (6)$$

where $p_{h1} = \frac{n_{11}}{n_{h1}}$, $p_{h2} = \frac{n_{12}}{n_{h2}}$, $\overline{p}_h = \frac{n_{h1}p_{h1} + n_{h2}p_{h2}}{n_{h.}}$, *and* $\overline{q}_h = 1 - \overline{p}_h$

Aside from the continuity correction in χ^2_{MH} and χ^2_C differ by a factor of $\left(\frac{n_{h..}}{n_{h..} - 1}\right)$ in each table. For moderate to large sample sizes per table, the difference between the two statistics is typically negligible. In general, though χ^2_{MH} offers advantages. Both statistics are asymptotically $\chi^2(1)$, but the quality of this approximation depends upon the table-specific sample sizes only for χ^2_C. In the extreme, χ^2_{MH} statistic will perform adequately in matched pair studies in which $n_{h..} = 2$ for all h, for sufficient total sample size. Cochran's statistic is not be used for such a situation. (Daniel et al, 2000, McDonald and Siebenaller, 1989).

In addition, the Mantel-Haenszel test has been shown to be optimal under the assumption of a constant odds ratio across tables (Birch, 1964) and it is asymptotically equivalent to likelihood ratio tests from unconditional and conditional logistic regression models for large strata and sparse data situations, respectively (Breslow and Day, 1980).

In addition to the test statistic, χ^2_{MH}, Mantel and Haenszel proposed an odds ratio estimator in their original 1959 paper. Their estimator is a weighted average of the table-specific observed odds ratios:

$$\hat{\psi}_{MH} = \frac{\sum_{h=1}^{H} v_h \hat{\psi}_h}{\sum_{h=1}^{H} v_h} = \frac{\sum_{h=1}^{H} R_h}{\sum_{h=1}^{H} S_h} \quad (7)$$

where $v_h = \frac{n_{h12}.n_{h21}}{n_{h..}}$, $\hat{\psi}_h = \frac{n_{h11}.n_{h22}}{n_{h12}.n_{h21}}$, $R_h = \frac{n_{h11}.n_{h22}}{n_{h..}}$, and $S_h = \frac{n_{h12}.n_{h21}}{n_{h..}}$

Since the introduction of χ^2_C, χ^2_{MH}, and $\hat{\psi}_{MH}$, a large literature has developed. Mantel-Haenszel-type estimators have been developed for the rate ratio (Rothman and Boice, 1979), rate difference (Greenland, 1982), risk ratio (Rothman and Boice, 1979; Tarone, 1981; Nurminen, 1981; and Kleinbaum, Kupper, and Morgenstern, 1982), and risk difference (Greenland, 1982).

Odds Ration in Mantel-Haenszel Method

The odds of an event (or condition) is defined by $\frac{\pi}{1-\pi}$, where π is the probability of the event. The odds ratio Ψ is the ratio of the odds of an event occurring in one group to the odds of that event in another group. In this research, these groups are sex (male and female) and students' grade (High or Low), or any other dichotomous classification. The odds ratio is used to test whether the probability of a certain event is the same for two groups. We note that the odds ratio takes values in (0,∞). An odds ratio of 1 indicates that the event under study is equally likely in both groups. If Ψ > 1, then the event is more likely in the first group, whereas Ψ < 1 indicates that it is less likely. The 2 × 2 table shows observations for two such groups and events A and $\bar{A}$, the complement of A.

	A	$\bar{A}$	Totals
Group 1	X_1	$n_1 - X_1$	n_1
Group 2	X_2	$n_2 - X_2$	n_2
Totals	$X_1 + X_2$	$n_1 + n_2 - X_1 - X_2$	$n_1 + n_2$

The odds ratio $\Psi = \dfrac{\pi_1 (1 - \pi_2)}{\pi_2 (1 - \pi_1)}$ is estimated by $\dfrac{X_1(n_2 - X_2)}{X_2(n_1 - X_1)}$, which is invariant if rows or columns (or both simultaneously) are interchanged. In clinical studies there are often only a few subjects. Multicentre trials increase the sample size, but populations differ for different centres and one cannot assume that probabilities for different centres are equal. However, one can assume that the odds ratios for each of the ***K*** centres are identical, that is, assuming a *common odds ratio* Ψ with $\Psi = \Psi_1 = \ldots = \Psi_k$. Under this *common odds ratio assumption*, the *Mantel-Haenszel (1959)* estimator $\hat{\Psi}$ of the common odds ratio is widely used by practising statisticians and epidemiologists. The MH estimator is a ratio of two sums C_{12} and C_{21}, where each summand of C_{ij} has the form $\dfrac{X_{ik}(n_{jk} - X_{jk})}{(n_{ik} + n_{jk})}$ with index ***k*** referring to the quantities of the ***k***th table or ***k***th centre. The factor $\dfrac{1}{(n_{ik} + n_{jk})}$ is a weight accounting for the sample size of the ***k***th table. The MH estimator is also often applied for other stratified data for which the common odds ratio assumption is reasonable.

Even if the assumption of a common odds ratio is slightly violated, the MH estimator is still a useful tool to summarise the association across tables. Despite the Mantel-Haenszel estimator's simplicity, it has some useful properties. First, it applies to very sparse data. More precisely, it is defined when only one summand of C_{12} and of C_{21} is non-zero (Suesse, 2009).

It is also *dually consistent*, that is, consistent under two types of asymptotic models: (i) when the sample size of each stratum increases and the number of strata is fixed, and (ii) when the number of observations becomes large with the number of strata, while the sample size of each stratum remains fixed. (i) is referred to as a *large-stratum* limiting model, or model (i), and to (ii) as a *sparse data* limiting model, or model (ii). In practice, model (i) represents large $n_{1k} + n_{2k}$ for each stratum and model (ii) represents large ***K***. The MH estimator is robust under any such extreme data. The consistency of the MH estimator for model (i) was shown by Gart (1962) and for model (ii) by Breslow (1981). Hauck (1979) derived the limiting variance of the MH estimator under model (i), whereas Breslow (1981) derived two asymptotic variances under model (ii): one based on the conditional distribution of the observations for each table given the marginal totals, and the other on the empirical variance. Applying either of the variance estimators depending on the given data, whether the data resembles the sparse data or large stratum case, is very unsatisfactory. Breslow and Liang (1982) proposed a weighted average of the two variance estimators to account for the two different limiting models. Robins, Breslow and Greenland (1986) proposed a variance estimator which is dually consistent under models (i) and (ii) based on the unconditional distribution of the data.

An alternative way to estimate the common odds ratio for $K \times 2 \times 2$ tables is to fit an ordinary logit model with main effects and no interaction, where the ***K*** strata and one binary classification are treated as factors and the other binary classification as a response. The corresponding loglinear model is a model with no three-way interaction among rows, columns and strata. However, the unconditional maximum likelihood (ML) estimator is a poor estimator, because under model II the nuisance parameters grow as the sample size grows. For instance when each table consists of a single matched pair, then the unconditional ML estimator of the common odds ratio converges to the square of the true common odds ratio (Anderson 1980, p.244). The nuisance parameters can be eliminated by conditioning on the margins of the 2×2 contingency table. The ML estimator based on the conditional distribution, which is non-central hypergeometric in each stratum, is also dually consistent. As a by-product, the ML fitting yields a variance estimator of the odds ratio estimator.

If the assumption of a common odds ratio fails, we can still use the MH estimate as a summary of the odds ratios among the strata. Without the common odds ratio assumption, the MH estimator is consistent under model (i) only; and appropriate standard errors were suggested by Guilbaud (1983), since the dually consistent variance estimator of Robins et al. (1986) fails.

A simple way to test the homogeneity of the odds ratio across strata is to apply a goodness-of-fit test to a logit model with only main effects and no interaction. The goodness-of-fit test statistic has $K - 1$ degrees of freedom (df) if the model holds.

Breslow-Day Test for Homogeneity of the Odds Ratios

Breslow and Day (1980) developed a test statistic which does not require model fitting and focuses directly on the potential lack of homogeneity. The Breslow-Day test statistic sums the squared deviations of observed and fitted values each standardised by its variance. According to Breslow and Day (1980) the test is used for stratified analysis of 2×2 tables to test the null hypothesis that the odds ratios for the *k*-strata are all equal. When the null hypothesis is true, the statistic has an asymptotic chi-square distribution with *k*-1 degrees of freedom.

Tarone (1985) proved that it is stochastically larger under the homogeneity assumption, and developed a modified score test statistic that is indeed asymptotically $\chi^2(K-1)$. Liang and Self (1985) proposed a score test assuming the log odds ratios across strata are independent and identically distributed, which is valid also when the sample size increases with the number of strata. Paul and Donner (1989) conducted a simulation study generally recommending Tarone's modified test statistic. Liu and Pierce (1993) used a different approach by assuming that the log odds ratios across the strata are a sample from a population with unknown mean and variance. They investigated the conditional likelihood functions for the mean and the variance. A test of homogeneity of the odds ratios can be conducted by testing whether the variance of the log odds ratio equals zero. Liu and Pierce (1993)'s approach is more general than that of Liang and Self (1985), since it describes the heterogeneity of the log odds ratios across the strata.
The estimation of the common odds ratio assumes that the strength of association as measured by the odds ratios in each sub-table is the same. This assumption is tested by the test of homogeneity of the odds ratio. To test this hypothesis, the Breslow-Day test is often employed. This statistic is compared to a standard X^2 distribution with $(H-1)$ degrees of freedom. The null hypothesis of homogeneity of odds ratio across the sub-tables is rejected if P-value is < α.
Breslow-Day test for stratified analysis of 2 ×2 tables tests the null hypothesis that the odds ratios for the *H*-strata are all equal. When the null hypothesis is true, the statistic has an asymptotic chi-square distribution with *k*-1 degrees of freedom.
The Breslow-Day statistic is computed as:

$$\hat{\theta}_{BD} = \frac{\sum_{h=1}^{H}\left(n_{h11} - E\left(n_{h11}\middle|\hat{\theta}_{MH}\right)\right)^2}{var\left(n_{h11}\middle|\hat{\theta}_{MH}\right)} \qquad (8)$$

where *E* and *var* denote expected value and variance, respectively. The summation does not include any table with a zero row or column.
It is advisable to test the homogeneity of the odds ratios in the different repeats, and if different repeats show significantly different odds ratios, the Cochran–Mantel–Haenszel test is not essential.

Estimating the Common Odds Ratio
While the Cochran-Mantel-Haenszel test provides the significance of the relationship between two variables (sex and final grade OR age at entry and final grade) across the sub-tables (department), it does not tell us the strength of this association. An estimator (The Mantel-Haenszel estimate of the common odds ratio in case-control studies) of the common odds ratio is given *by:*

$$\hat{\theta}_{MH} = \frac{\sum_{h=1}^{H}\left(\frac{n_{h11} * n_{h22}}{n_{h..}}\right)}{\sum_{h=1}^{H}\left(\frac{n_{h12} * n_{h21}}{n_{h..}}\right)} \qquad (9)$$

If the confidence interval of the common odds does not include 1; then, we conclude that there is dependence on the two variables of interest across the sub-tables else.
The estimate of the common odds ratio is based on the assumption that the strength of the association is the same in each department. If this were not the case, then we would have believed that there is *interaction* or *effect modification* between department and final grade. The factor variable (department) is often referred to as the *effect modifier.* Evidence of the homogeneity of the odds ratios across departments indicates that there is no significant effect modification in this case.

Using the estimated variance for log($\hat{\hat{\theta}}_{MH}$) given by Robins, Breslow, and Greenland (1986), we can compute the corresponding 100(1 – α)%% confidence limits for the odds ratio as:

$$\left(\hat{\theta}_{MH} * \exp(-z\hat{\sigma}), \hat{\theta}_{MH} * \exp(z\hat{\sigma})\right) \qquad (10)$$

where

$$\hat{\sigma}^2 = var\left[\ln\hat{\theta}_{MH}\right] = \frac{\frac{\sum_{h=1}^{H}(n_{h11}+n_{h22})(n_{h11}n_{h22})}{n_{h..}^2}}{2\left(\frac{\sum_{h=1}^{H}n_{h11}n_{h22}}{n_{h..}}\right)^2} + \frac{\frac{\sum_{h=1}^{H}(n_{h12}+n_{h21})(n_{h12}n_{h21})}{n_{h..}^2}}{2\left(\frac{\sum_{h=1}^{H}n_{h12}n_{h21}}{n_{h..}}\right)^2}$$

$$+\frac{\frac{\sum_{h=1}^{H}(n_{h11}+n_{h22})(n_{h12}n_{h21})+(n_{h12}+n_{h21})(n_{h11}n_{h22})}{n_{h..}^{2}}}{2\left(\frac{\sum_{h=1}^{H}n_{h11}n_{h22}}{n_{h..}}\right)\left(\frac{\sum_{h=1}^{H}n_{h12}n_{h21}}{n_{h..}}\right)}$$

Analysis

The table below shows the cumulative odds of 623 graduated students of Faculty of Science, University of Ilorin, Nigeria for 2011/2012 academic session.

Table 1 indicates that the odds for students graduating with first class against all other grades is 0.0196. The odds for students graduating with second class upper and above against second class lower and below is 0.5270 while it is 3.5145 for students graduating with second class lower and above against third class and pass. Also, the odds for students graduating with third class and above against graduating with pass is 68.2222

The Odds of graduating with second class upper and above (0.5270) is about half of graduating with second class lower and below. Though, this odds favours students graduating with second class lower and below, it is reasonable to infer that the final grade is approximately symmetrical about two groups. The first group are those with Second Class Lower and below (Low Grade) while the other is for those with Second Class Upper and above (High Grade). Also, in most of researches involving categorising students grade in Nigeria varsities, First Class and Second Class Upper are usually grouped together while other grades are grouped together. Hence, the justification to categorize the final grade in to High Grade (Second Class Upper and First Class) and Low Grade (Second Class Lower, Third Class, and Pass) for Cochran's and Mantel-Haenszel statistics. There is no CO for Pass or above since this includes the whole sample, i.e. the CP is 1 (or 100%).

In Table 2, both Breslow-Day and Tarone's statistics show that the null hypothesis of homogeneity of odds ratio across the departments is not rejected since P-value (0.861) > α (0.05). This implies that the odds ratio across the ten departments (relating to age at entry and final grade) are all equal.

Hypothesis Statement

H_0: Controlling for (or within departments), there is no relationship between age at entry and final grade (Low Grade or High Grade).

H_a: Controlling for (or within departments), there is relationship between age at entry and final grade (Low Grade or High Grade).

From Table 3, the P-values (0.433 and 0.493) for both Cochran's and Mantel-Haenszel statistics reveals that the null hypothesis of no association of final grade and age at entry of students among the ten departments of faculty of science, University of Ilorin is not rejected. Hence, we conclude that the final grade of students (Low Grade or High Grade) is not associated with age students at entry.

In Table 4, the odds in favour of a student whose age is less than 20 years graduating with Low Grade (Pass, Third Class, and Second Class Lower) is 0.865. The 95% confidence interval for this common odds ratio is (0.601, 1.246). This interval includes 1; therefore, age at entry of students is independence on final grade in two categories (High Grade or Low Grade).

From Table 5, both Breslow-Day and Tarone's statistics show that the null hypothesis of homogeneity of odds ratio across the departments is not rejected since P-value (0.661) > α (0.05). This implies that the odds ratio across the ten departments (relating to sex and final grade) are all equal.

Hypothesis Statement

H_0: Controlling for (or within departments), there is no relationship between gender and final grade (Low Grade or High Grade).

H_a: Controlling for (or within departments), there is relationship between gender and final grade (Low Grade or High Grade).

In Table 6 also, the P-values (0.031 and 0.040) for both Cochran's and Mantel-Haenszel statistics reveals that the null hypothesis of no association of final grade and sex of students among the ten departments of faculty of science, University of Ilorin is not rejected. Hence, we conclude that the final grade of students (High Grade or Low Grade) is not associated with sex of students.

From Table 7, the odds in favour of male students graduating with Low Grade (Pass, Third Class, and Second Class Lower) is 0.670. The 95% confidence interval for this common odds ratio is (0.465, 0.966). This interval does not include 1; therefore, there is dependence on sex of the students and final grade in two categories (High Grade or Low Grade).

Conclusion

From the analysis carried out in this research, it can be concluded that the odds of graduating with Second Class Upper and above is about half of graduating with Second Class Lower and below. The odds ratio across the ten departments (relating to age at entry & final grade and sex & final grade) are all equal and the final grade of students (Low Grade or High Grade) is not associated with both sex and age students at entry. Also, the odds in favour of a student whose age is less than 20 years graduating with Low Grade is 0.865 while it is 0.670 for male students graduating with lower grade.

References

Anderson, E. B. (1980): *Discrete Statistical Models with Social Science Applications*, North-Holland, New York.
Bickel, P. J., and O'Connell, J. W. (1975): *Is there a sex bias in graduate admissions? Science, 187,* 398–404.
Birch, M. W. (1964): *The detection of partial association: the 2 X 2 case,* Journal of the Royal Statistical Society, B 26, 313 – 324
Breslow, N. E. (1981): *Odds ratio estimators when the data are sparse*, Biometrika 68(1), 73–84.
Breslow, N. and Day, N. (1980): *Statistical Methods in Cancer Research I: The Analysis of Case-Control Studies*, International Agency for Research on Cancer, Lyon.
Breslow, N. E. and Liang, K. Y. (1982): *The variance of the Mantel-Haenszel estimator,* Biometrics 38(4), 943–952.
Cochran, W. G. (1954): *Some methods for strengthening the common 2x2 tests*, Biometrics 10(4), 417–451.
Daniel B. Hall, Robert F. Woolson, William R. Clarke and Martha F. Jones (2000): *Cochran_Mantel_Haenszel Techniques: Applications Involving Epidemiologic Survey Data, Bioenvironmental and Public Health Statistics, Handbook of Statistics, vol(18), pp 483 – 500*
Gart, J. J. (1962): *On combination of relative risks*, Biometrics 18(4), 594–600.
Greenland, S. (1982): Interpretation and estimation of summary ratios under heterogeneity, Statistics in Medicine, (1), 217 – 227
Guilbaud, O. (1983): *On the large-sample distribution of the Mantel-Haenszel odds-ratio estimator,* Biometrics 39(2), 523–525.
Hauck, W. W. (1979): *Large sample variance of the Mantel-Haenszel estimator of a common odds ratio*, Biometrics 35(4), 817–819.
Kleinbaum, D. G., Kupper. L. L. and Morgenstern. H. (1982): *Epidemiologic Research: Principles and Quantitative Methods*, Lifetime Learning Publications, Belmont, CA.
Lawal Bayo (2003): *Categorical Data Analysis with SAS and SPSS Applications,* Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersy. p. 138
Liang, K. Y. and Self, S. G. (1985): *Tests for homogeneity of odds ratio when the data are sparse*, Biometrika 72(2), 353–358.
Liu, Q. and Pierce, D. A. (1993): *Heterogeneity in Mantel-Haenszel-type models,* Biometrika 80(3), 543–556.
Mantel, N. and Haenszel, W. (1959): *Statistical aspects of the analysis of data from retrospective studies of disease*, Journal of the National Cancer Institute 22, 719–748.
McDonald, J.H. and J.F. Siebenaller (1989): *Similar geographic variation at the Lap locus in the mussels Mytilus trossulus and M. edulis.* Evolution 43: 228-231.
Nurminen, M. (1981): Asymptotic efficiency of general non-iterative estimators of common relative risk, Biometrika (68), 525 – 530
Paul, S. R. and Donner, A. (1989): *A comparison of tests of homogeneity of odds ratios in Kx2x2 tables*, Statistics in Medicine 8(12), 1455–1468.
Robins, J., Breslow, N. and Greenland, S. (1986): *Estimators of the Mantel-Haenszel variance consistent in both sparse data and large-strata limiting models*, Biometrics 42(2), 311–323.
Rothman, K. J. and Boice, J. D. (1979): Epidemiologic Analysis with a Programmable Calculator, U.S. Government Printing Office, Washington, DC.
Suesse, Thomas Falk (2009): *Analysis and Diagnostics of Categorical Variables with Multiple Outcomes,* Unpublished PhD thesis, Victoria University of Wellington
Tarone, R. E. (1981): On summary estimators of relative risk, Journal of Chronic Disease (34), 463 – 468
Tarone, R. E. (1985): *On heterogeneity tests based on efficient scores*, Biometrika 72(1), 91–95.

Appendix

Table 1: Cumulative Odds for Final Grade

Final Grade	Pass	Third Class	Second Class Lower	Second Class Upper	First Class
Frequency	9	129	270	203	12
CS	623	614	485	215	12
CP	1.0000	0.9856	0.7785	0.3451	0.0193
CO		68.2222	3.5145	0.5270	0.0196

CS = Cumulative Sum CP = Cumulative Proportion CO = Cumulative Odds

$$CP = \frac{CS}{623} \qquad CO = \frac{CS}{623 - CS} = \frac{CP}{1 - CP}$$

Table 2: Tests of Homogeneity of the Odds Ratio (age)(Age at entry * Final Grade CMH * Department Analysis)

	Chi-Squared	df	Asymp. Sig. (2-sided)
Breslow-Day	4.678	9	.861
Tarone's	4.677	9	.861

Table 3: Tests of Conditional Independence (age)

	Chi-Squared	df	Asymp. Sig. (2-sided)
Cochran's	.614	1	.433
Mantel-Haenszel	.469	1	.493

Table 4: Mantel-Haenszel Common Odds Ratio Estimate (age)

Estimate			.865
ln(Estimate)			-.145
Std. Error of ln(Estimate)			.186
Asymp. Sig. (2-sided)			.437
Asymp. 95% Confidence Interval	Common Odds Ratio	Lower Bound	.601
		Upper Bound	1.246
	ln(Common Odds Ratio)	Lower Bound	-.509
		Upper Bound	.220

Table 5: Tests of Homogeneity of the Odds Ratio (sex)(Sex * Final Grade CMH * Department Analysis)

	Chi-Squared	df	Asymp. Sig. (2-sided)
Breslow-Day	6.767	9	.661
Tarone's	6.767	9	.661

Table 6: Tests of Conditional Independence (sex)

	Chi-Squared	df	Asymp. Sig. (2-sided)
Cochran's	4.671	1	.031
Mantel-Haenszel	4.214	1	.040

Table 7: Mantel-Haenszel Common Odds Ratio Estimate (sex)

Estimate			.670
ln(Estimate)			-.401
Std. Error of ln(Estimate)			.187
Asymp. Sig. (2-sided)			.032
Asymp. 95% Confidence Interval	Common Odds Ratio	Lower Bound	.465
		Upper Bound	.966
	ln(Common Odds Ratio)	Lower Bound	-.767
		Upper Bound	-.035