ORİJİNAL ARAŞTIRMA ORIGINAL RESEARCH

Closed Form Methods to Compare Two Independent Proportions for Clustered Data

Kümelenmiş Verilerde Bağımsız İki Oranın Karşılaştırılmasında Kullanılan İteratif Olmayan Yöntemler

Derya ÖZTUNA,[a]
Yasemin GENÇ[a]

[a]Department of Biostatistics,
Ankara University Faculty of Medicine,
Ankara

Geliş Tarihi/*Received:* 28.05.2009
Kabul Tarihi/*Accepted:* 22.10.2009

Bu makalede yer alan bilgilerin ki-kare testi ile ilgili bölümü, 8-12 Mayıs 2005 tarihinde Antalya'da düzenlenen 4. İstatistik Kongresi'nde sözlü bildiri olarak sunulmuştur.

Yazışma Adresi/*Correspondence:*
Derya ÖZTUNA
Ankara University Faculty of Medicine,
Department of Biostatistics,
Ankara,
TÜRKİYE/TURKEY

ABSTRACT Objective: Many medical researches involve the collection of multiple observations for each subject studied. The statistical literature typically refers to such data as "clustered data". In order to compare two proportions estimated from independent samples, chi-squared test can be used. However, this test is not appropriate for clustered data since it doesn't take the dependencies among observations within the same cluster into account. According to simulation studies, when the within-correlation coefficient is positive, ignoring the correlation results in inflation to Type I error rate. So, methods taking the within-cluster correlation into account should be used to perform chi-square test in clustered data. **Material and Methods:** In this study, closed form methods for chi-square test to compare the proportion of decayed tooth estimated from independent samples (public and private school) and Mantel-Haenzel chi-square test supposing that age can be a confounder for the proportion of decayed tooth, will be introduced and applied on an ophthalmology data. **Results:** According to the results, although there were significant differences between the proportion of decayed tooth among the selected public and private schools at standard chi-squared test (p=0.047), this difference was not significant at methods which was adjusted to the clustered binary data (p>0.05). When the Mantel-Haenszel chi-square test adjusted for age was applied, the p value associated with standard Mantel-Haenszel test was 0.06 and those for the adjusted methods were 0.497 and 0.351. **Conclusion:** When comparing two independent proportions for clustered binary data, methods taking the within-cluster correlation into account should be used.

Key Words: Clustered data, intracluster correlation, chi-squared test, mantel-haenszel test

ÖZET Amaç: Tıp alanında yapılan çalışmaların çoğunda, bir denekten birden fazla ölçüm alınır. İstatistik literatüründe bu tür veriler "kümelenmiş veri" olarak adlandırılır. Bağımsız gruplardan tahmin edilen iki oranın karşılaştırılmasında ki-kare testi kullanılabilir. Ancak bu test, bir kümedeki ölçümler arasındaki bağımlılık yapısını hesaba katmadığından kümelenmiş veriler için uygun değildir. Yapılan benzetim çalışmaları, bağımlılık varsayımının bozulmasının, küme içi pozitif korelasyon varlığında I. tip hata oranında artışa neden olduğunu göstermiştir. Bu nedenle kümelenmiş verilerde ki-kare testinin uygulanmasında küme-içi korelasyonu dikkate alan yöntemler kullanılmalıdır. **Gereç ve Yöntemler:** Çalışmamızda, kümelenmiş verilerde bağımsız örneklemlerden (devlet ve özel okul) elde edilen çürük diş oranlarının karşılaştırılmasında ki-kare testi ve yaş değişkeninin çürük diş oranları üzerinde karıştırıcı etkisinin olabileceği düşüncesi ile Mantel-Haenszel testi için iteratif olmayan yöntemler tanıtılmış ve diş hekimliği alanından bir veri seti üzerinde uygulanmıştır. **Bulgular:** Elde edilen sonuçlar standart ki-kare testinden elde edilenle karşılaştırıldığında standart ki-kare testinde, seçilen devlet ve özel okul için çürük diş oranları arasında anlamlı farklılık bulunurken (p=0.047), kümelenmiş veriler için geliştirilmiş yöntemlerde bu farklılık ortadan kalkmıştır (tüm yöntemler için p>0.05). Yaş değişkeni karıştırıcı değişken olarak alınarak Mantel-Haenszel testi uygulandığında, standart Mantel-Haenszel testi için p değeri 0.06 iken, düzeltilmiş yöntemler için p değerleri 0.497 ve 0.351 olarak tespit edilmiştir. **Sonuç:** Kümelenmiş verilerde bağımsız iki oranın karşılaştırılmasında, küme içi korelasyonu dikkate alan yöntemler kullanılmalıdır.

Anahtar Kelimeler: Kümelenmiş veri, küme içi korelasyon, ki-kare testi, mantel-haenszel testi

Turkiye Klinikleri J Biostat 2010;2(1):11-7

The most common problem encountered in many medical studies is the incorrect statistical analysis of multiple observations taken from the same subject. While observations from the different subjects can be considered statistically independent, different observations from the same subject are correlated. The actual correlation varies from study to study and measurement to measurement. This data structure is called as "clustered data" in statistical literature.

Clustered data occur frequently in several fields of studies. For example, in periodontal studies, observations can be taken from multiple sites (gums, teeth or tooth surfaces) on each subject. In ophthalmologic studies, the subject again is the cluster, but two eyes are measured, for example; the presence of uveitis can be exposed by considering either or both eyes of a patient with Behçet's disease. In teratologic studies the litter is the cluster, but measurements are taken from each animal in the litter, for example; in comparison of the survival rate of animals in a treatment group to the corresponding rate in a control group, each animal in the litter is considered. A special class of the studies is community intervention trials in which medical practices, factories, or entire cities are taken as cluster.

Conventional statistical methods are not appropriate for clustered data, since they do not take into account the dependencies among observations within the same cluster. Therefore, several statistical methods have been proposed in the literature. Some of these are based on likelihood inference and called "likelihood ratio (LR) methods". They require parametric model assumptions for the intracluster correlation -for example beta-binomial model,[1-3] the correlated-binomial model[4], the multiplicative-binomial model,[5] and the correlated probit regression model.[6]

Liang and Zeger[7] introduced a class of generalized estimating equations (GEE), based on moment methods that provide consistent estimates of regression parameters and their variances in the context of generalized linear models making no distributional assumptions. However, to make reliable estimates, these methods require fairly large sample sizes (more than 40 clusters per group). Another shortcoming of them is the requirement for computationally intensive iterative solution.

In our study, closed form methods for comparing two proportions and Mantel-Haenzel chi-square test for comparing two proportions estimated from independent observations in a series of 2x2 tables of clustered binary data will be introduced with illustrative applications.

METHODS COMPARED

We restrict our investigation in this paper to the comparison of proportions in two groups, but extensions to more than two groups are straightforward. The main notations for the methods discussed below are given in Table 1.

We further define "overall event-rate" as $\hat{P} = \sum_{i=1}^{2} x_i / \sum_{i=1}^{2} n_i$, "average cluster size" as $\bar{n}_i = \sum_{j=1}^{m_i} n_{ij} / m_i$, total number of cluster as $M = \sum_{i=1}^{2} m_i$ and total number of observations as $N = \sum_{i=1}^{2} n_i$.

TABLE 1: Main notations.

	Group i (i=1,2)		
	Number of event	**Number of observation**	**Cluster-specific event rate**
Cluster j (j=1,...,m_i)	**(x_{ij})**	**(n_{ij})**	$\hat{p}_{ij} = x_{ij} / n_{ij}$
1	x_{i1}	n_{i1}	$\hat{p}_{i1}$
2	x_{i2}	n_{i2}	$\hat{p}_{i2}$
...	...	...	...
m_i	x_{im_i}	n_{im_i}	$\hat{p}_{im_i}$
Total	x_i	n_i	$\hat{p}_i = x_i / n_i$

METHODS FOR SINGLE 2X2 TABLES

Method Based on Direct Adjustment of Pearson Chi-Squared Statistic

Extending the Brier method,[8] Donner&Banting[9] and Donner[10] proposed an adjustment for Pearson chi-squared statistic which depends on clustering effects computed seperately in each treatment group. Following a one degree of freedom chi-square distribution, this statistic is given by

$$\chi_A^2 = \sum_{i=1}^{2} \left\{ \frac{(x_i - n_i\hat{P})^2}{n_i C_i \hat{P}(1-\hat{P})} \right\}, \qquad (1)$$

where $C_i = 1 + (\bar{n}_{Ai} - 1)\hat{\rho}$ is an estimate of the clustering effect in group i, i=1,2. Here, $\bar{n}_{Ai}$ is the second moment for cluster sizes.

The estimator $\hat{\rho}$ referred to as the "analysis of variance" estimator for intracluster correlation coefficient ρ is given by

$$\hat{\rho} = \frac{MSC - MSE}{MSC + (n_0 - 1)MSE} \qquad (2)$$

$$MSC = \sum_{i=1}^{2} \sum_{j=1}^{m_i} n_{ij}(\hat{p}_{ij} - \hat{p}_i)^2 / (M-2)$$

$$MSE = \sum_{i=1}^{2} \sum_{j=1}^{m_i} n_{ij}\hat{p}_{ij}(1 - \hat{p}_{ij}) / (N-M)$$

$$n_0 = \left[N - \sum_{i=1}^{2} \bar{n}_{Ai} \right] / (M-2), \qquad \bar{n}_{Ai} = \sum_{j=1}^{m_i} n_{ij}^2 / n_i$$

The mean square errors MSC and MSE measure the variation in response among and within clusters, respectively. If responses within a cluster are no more similar than responses within different clusters, then MSC=MSE, $\hat{\rho}$ =0 and C_i=1, so χ^2_A reduces to the standard Pearson chi-squared test statistic. If all responses within a cluster are identically same, then MSE=0 and $\hat{\rho}$ =1.

This method assumes that clustering effects are homogeneous across groups. Under the assumption that the intracluster correlation coefficient is positive, p value computed from the standard Pearson chi-squared test is likely to be biased downward and so the difference between proportions may not be statistically significant at the 5 percent level. The magnitude of the bias associated with the standard Pearson chi-squared test statistic increases with both the value of $\hat{\rho}$ and $\bar{n}_{Ai}$.

Methods Based on Ratio Estimate Theory

Rao and Scott[11] proposed a method for testing null hypothesis that regards the observed event rate $\hat{p}_i$ as ratios rather than proportions. However, the estimated clustering effects in each group are developed quite differently. If observations rather than clusters were the unit of allocation, then the estimated variance of $\hat{p}_i$ would be given by the usual binomial expression $Var_B(\hat{p}_i) = \hat{p}_i(1-\hat{p}_i)/n_i$. However $Var_B(\hat{p}_i)$ tends to underestimate the true variance of $\hat{p}_i$ if the unit of allocation is an intact cluster. If $\hat{p}_i$ is regarded as a ratio, an appropriate variance estimate is obtained from standard sample survey theory[12] as $Var_R(\hat{p}_i) = m_i(m_i-1)^{-1} n_i^{-2} \sum_{j}^{m_i} r_{ij}^2$ where $r_{ij} = x_{ij} - n_{ij}\hat{p}_i$. The estimated clustering effects in each grup are defined as $d_i = Var_R(\hat{p}_i)/Var_B(\hat{p}_i)$. Let "effective sample size in each group" $\tilde{n}_i = n_i/d_i$ "the effective number of event" $\tilde{x}_i = x_i/d_i$, "effective event rate" $\tilde{p}_i = \sum_{i=1}^{2} \tilde{x}_i/\tilde{n}_i$ and "the overall event-rate" $\tilde{P} = \sum_{i=1}^{2} \tilde{x}_i / \sum_{i=1}^{2} \tilde{n}_i$. The resulting test statistic is obtained by substituting these quantities in the standard Pearson chi-squared statistic,

$$\chi_R^2 = \sum_{i=1}^{2} (\tilde{x}_i - \tilde{n}_i\tilde{P})^2 / \left[\tilde{n}_i\tilde{P}(1-\tilde{P})\right] \qquad (3)$$

The statistic makes no assumptions concerning the nature of the clustering, and, in particular, does not assume that the population clustering effects in the two comparison groups are equal. However, the number of clusters in each group must be large in order to ensure that χ^2_R follows a one degree of freedom chi-square distribution under H_0.

Rao and Scott also proposed a version of χ^2_R which depends on a pooled estimate of an assumed common clustering effect. The resulting statistic is given by $\chi^2_{RP=}\chi^2_P/d$ where

$$d = \sum_{i=1}^{2} \frac{(1 - n_i/M)\hat{p}_i(1-\hat{p}_i)d_i}{\hat{P}(1-\hat{P})}$$

Other Simple Methods

Another simple method is to apply the "standard two-sample t test" to the clustered binary data. In this method using event rates for each cluster, mean event rates are obtained and compared. Test statistic yields a t distribution with a (m_1+m_2-2) degree of freedom.

The main assumption of this method is that cluster-specific event rates are normally distributed with homogeneous variance. However, a theoretical objection that may be raised against this method is that this assumption will not be true, if cluster sizes are fairly different. Nevertheless, simulation research has shown that t test is remarkably robust to violations of the underlying assumptions and the resulting P values are likely to be accurate to a reasonable approximation.[13] As another objection, this method completely ignores any variation in cluster size, giving each of the observed event rates equal weight. Not being able to make inference about odds ratio is final objection of two-sample t test.

In case of the lack of applicability of two sample t test, a nonparametric approach; Wilcoxon rank-sum test may be used. In this method, two samples are pooled and the cluster-specific event rates are ranked by size. This method is valid even in small samples. However, the rank-sum test ignores not only the variation in cluster size, but also the actual magnitudes of event rates. Also, if the number of clusters per group is less than four, achieving statistical significance at $p< 0.05$ is impossible for two-tailed tests, regardless of the magnitude of treatment effects.

METHODS FOR MULTIPLE 2X2 TABLES

In this section, we focus on analyzing data that can be summarized in a series of 2x2 tables. In such an analysis, we are concerned with detecting an association between a binary response measurement and a binary exposure, stratified across levels of a third, categorical covariate. To describe the strength of the association between response and exposure, it is common to report the odds ratio (OR).

For clustered binary data, Donald and Donner,[14] Donner and Banting,[9] Rao and Scott,[11] Zhang and Boos[15] and Liang[16] have developed simple corrections to standard Mantel-Haenszel test.[17,18] These approaches are limited, however, because they assume that the exposure and covariate information collected does not vary from observation to observation. In contrast, regression-type approaches such as GEE method[7] or random-effects approaches[19-21] can handle observation-specific covariates. While more flexible, these complex methods may be difficult to apply in practice due to computational problems. As an alternative of these complex methods, Begg[22] and Begg and Panageas[23] proposed a closed-form method whether the covariates are cluster-specific or observation-specific.[24]

If we consider a series of K 2x2 tables with fixed row totals (n_{1k}, n_{2k}), number of event (x_{1k}, x_{2k}) and number of non-event (y_{1k}, y_{2k}) with associated event rates (p_{1k}, p_{2k}), k=1, … , K. It is of interest to test the null hypothesis $H_0{:}\psi=1$, assuming a common odds ratio $\psi= (p_{1k}q_{2k})/(p_{2k}q_{1k})$ where $q_{ik}=1-p_{ik}$, i=1,2. Under the assumption that x_{ik}'s are independent binomial variables $B(n_{ik}, p_{ik})$, well-known Mantel-Haenszel chi-square statistic is given by:

$$\chi^2_{MH} = \frac{\left[\sum_{k=1}^{K} \left(\frac{x_{1k}y_{2k} - x_{2k}y_{1k}}{n_k}\right)\right]^2}{\sum_{k=1}^{K} \frac{n_{1k}n_{2k}x_k y_k}{(n_k-1)n_k^2}} \qquad (4)$$

where $x_k= (x_{1k} + x_{2k})$, $y_k= (y_{1k} + y_{2k})$, $n_k= n_{1k} + n_{2k}$. χ^2_{MH} is asymtotically (as $n_{ik}\rightarrow\infty$ for each i, k) χ^2 with 1 degree of freedom.

In this paper, because of their simplicity, we will introduce "Rao and Scott" and "Donald and Donner" approaches.

Rao and Scott approach is based on the clustering effects for each group (d_i), mentioned in Section 2.1.2. According to this approach, adjusted Mantel-Haenszel chi-square statistic, $\chi^2_{MH(R)}$, as χ^2 with one degree of freedom, is now obtained by replacing (x_{ik}, n_{ik}) with $(\tilde{x}_{ik}, \tilde{n}_{ik})$ in (4), where $\tilde{x}_{ik} = x_{ik}/d_{ik}$ and $\tilde{n}_{ik} = n_{ik}/d_{ik}$.

In Donald and Donner approach, the correction factor C_n to adjust sample sizes implicit in the Mantel-Haenszel statistic is used. When cluster size (n) is constant, the adjusted factor C_n is defined as $1+(n-1)\rho$. Donald and Donner also provided formulae of the adjusted chi-square test in the imbalanced setting where c luster size varies. The correction factor for group i of stratum k is the weighted mean in the chi-square statistic:

$$B_{ik} = \frac{\sum_n n m_{ikn} C_n}{\sum_n n m_{ikn}}$$

$$\chi^2_{MH(A)} = \frac{\left[\sum_{k=1}^{K} \left(\frac{n_{1k}y_{2k} - n_{2k}y_{1k}}{n_{1k}D_{1k} + n_{2k}D_{2k}}\right)\right]^2}{\sum_{k=1}^{K} \frac{n_{1k}n_{2k}\hat{P}_k\hat{Q}_k}{(n_{1k}D_{2k} + n_{2k}D_{1k} - 1)n_k^2}} \quad (5)$$

where $C_n = 1 + (n-1)\hat{\rho}_{ik}$, $\hat{\rho}_{ik} = 1 - \frac{\sum_{j=1}^{m_{ik}} y_{ijk}(n_{ijk} - y_{ijk})/n_{ijk}}{m_{ik}(n_{ik} - 1)\hat{p}_{ik}(1 - \hat{p}_{ik})}$ and D_{ink} (n=1, … n) denotes the number of clusters of size n in group i of stratum k.

AN EXAMPLE

As an example, we consider data arising from a periodontal study conducted by Tulunoğlu et al.[25] In this experiment, a public and a private school in Ankara were randomly selected to compare the proportion of decayed tooth. The aim of the investigation is to test whether the proportion of decayed tooth among public school $\hat{P}_1 = 199/1735 = 0.115$ is significantly different from the corresponding proportion among private school $\hat{P}_2 = 191/2004 = 0.095$.

The Pearson chi-square statistic is inappropriate for these data, since the responses of teeth within a mouth are more likely to the similar than responses of teeth in different mouth. The magnitude of within-mouth correlation is $\hat{\rho} = 0.121$.

In our study, the number of cluster is large for each school (m_1=78; m_2=92), degree of variation in cluster size is small (for public school the range is (16,24) and for private school the range is (17,24) and degree of intracluster correlation are relatively small for each school ($\hat{\rho}_1 = 0.116$, $\hat{\rho}_2 = 0.127$). Since the type of our study is an observational study and the ratio estimate reaches its optimal performance under the conditions which are hold in our study, usage of ratio estimate chi-square test is more appropriate than other methods we discussed. The values of all test statistics discussed above are presented in Table 2 whether they are appropriate or not.

It is interesting to note that the value of Pearson chi-square is the only one that shows at 5% statistical significance levels. The P values for other methods vary between 0.108 and 0.294 indicating that there is no difference between the proportions of decayed tooth for public and private school.

Although the distiribution of the proportion of decayed tooth for each school was not normal, the P value obtained from two-sample t test was very close to that of other methods. This property of two-sample t test which was pointed out in other studies was proven in our study also.

Starting out the point that age could be a confounder for the difference between the proportions, we applied Mantel-Haenszel chi-square test adjusted for age. The data are reported in Table 3, classified by school (public school denoted by i=1 and private school by i=2) and age {age=7 (k=1), age=8 (k=2)}.

The number of mouth, m_{ik}, varied from 30 to 62. We computed clustering effect as d_{11}=3.896, d_{12}=2.841 for public school and d_{21}=3.612, d_{22}=3.853 for private school. We computed also intracluster correlation coefficients as ρ_{11}=0.142, ρ_{12}=0.079 for public school and ρ_{21}=0.128, ρ_{22}=0.119 for private school.

The question of interest is whether the proportion of decayed tooth among public school is different from the corresponding proportion among private school. We obtain the observed values of χ^2_{MH}, $\chi^2_{MH(R)}$ and $\chi^2_{MH(A)}$ as 3.58, 0.462 and 0.87 respectively. While the P value associated with

TABLE 2: Summary of results.

Method	Test statistic	P
Standard Pearson chi-square test	χ^2_{P}= 3.945	0.047
Adjusted chi-square test	χ^2_{A}= 1.100	0.294
Ratio estimate chi-square test	χ^2_{R}= 1.111	0.292
Pooled ratio estimate chi-square test	χ^2_{RP}= 1.106	0.293
Two-sample t test	t=1.251	0.213
Nonparametric approach- Wilcoxon rank-sum test	z=1.608	0.108

TABLE 3: Summary of data for Mantel-Haenszel test.

Age	School	No.of mouth	No.of decayed tooth	Total no. of tooth	Proportion of decayed tooth	Odds ratio
Age=7	Public	43	111	922	0.120	1.47
	Private	62	111	1303	0.085	
Age=8	Public	35	88	813	0.108	0.96
	Private	30	80	711	0.113	

standard Mantel-Haenszel statistic is 0.06, 0.497 for Rao-Scott MH statistic and 0.351 for Donald-Donner MH statistic. Although all statistics indicate non-significant school difference, effect of adjustment on P values is striking.

DISCUSSION

The aim of many researches is to compare the event rates of individuals in each of several groups that have a certain characteristics. In our study, we have presented closed-form methods for clustered-binary data to compare two event rates estimated from independent groups. These methods include adjusted chi-square test, ratio estimate chi-square test, pooled-ratio estimate chi-square test, two-sample t test and Wilcoxon rank-sum test-non-parametric approach for single 2x2 tables and Mantel-Haenszel tests for multiple 2x2 tables.

The different areas of usage for these methods will be discussed under the light of results in different studies.

It is recommended on that the choice of method should depend on whether the study is experimental or observational. Experimental studies in which the assignment of clusters to different groups is random, assure the validity of assumption of a common population design effect. So, adjusted and pooled ratio estimate chi-square tests which require this assumption are appropriate for these studies. Brier[8] points out that these tests provide adequate results even if the design effects are moderately different.

However, for observational studies, assumption of a common population design effect doesn't hold generally. In this situation, usage of ratio estimate chi-square test, GEE and LR methods are recommended. Among the closed form methods, ratio estimate chi-square test is well suited to non-randomized studies, particularly those involving systematic differences in mean cluster size from group to group. This method reaches its optimal performance when the number of cluster is large; degree of variation in cluster size and degree of intracluster correlation are relatively small for each group.[12] Also, adjusted chi-square test can be used in observational studies under the assumption of equal intracluster correlation coefficients. When this assumption is violated, this test is recommended only when the two groups have an equal number of clusters. Evidence exists that the different cluster size distributions do not seriously affect the validity of the adjusted chi-square test.[26] As few as 10 clusters per group is enough to ensure the validity of adjusted chi-square test and this can be thought as an advantage of this test compared to ratio estimate chi-square test.

A particular class of the studies is community intervention trials, in which the number of cluster per group is too small (often 10 or less), but the number of observation is relatively large, such as schools, factories, or entire communities. Because of these properties of such studies, only two-sample t test may be applied safely with as few as three clusters per group even though individual event rates are not likely to be normally distributed. Also, the non-parametric procedure can be used if the number of clusters per group is at least four. The principal disadvantage of the t test relative to other procedures is that it does not reduce to standard methods for testing H_0: $P_1=P_2$ in the absence of clustering. It doesn't yield results more interpretable on an odds ratio scale while other procedures, such as the adjusted chi-square tests and the GEE approach do.

REFERENCES

1. Williams DA. The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. Biometrics 1975;31(4):949-52.
2. Crowder MJ. Beta-binomial ANOVA for proportions. Appl Stat 1978;27(1):34-7.
3. Paul SR. Analysis of proportions of affected foetuses in teratological experiments. Biometrics 1982;38(2):361-70.
4. Kupper LL, Haseman JK. The use of a correlated binomial model for the analysis of certain toxicological experiments. Biometrics 1978;34(1):69-76.
5. Altham PME. Two generalizations of the binomial distribution. Appl Statist 1978;27(3): 162-7.
6. Ochi Y, Prentice RL. Likelihood inference in a correlated probit regression model. Biometrika 1984;71(3):531-43.
7. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;73(1):13-22.
8. Brier SS. Analysis of contingency tables under cluster sampling. Biometrika 1980;67(3):591-6.
9. Donner A, Banting D. Analysis of site-specific data in dental studies. J Dent Res 1988;67(11):1392-5.
10. Donner A. Statistical methods in ophthalmology: an adjusted chi-square approach. Biometrics 1989;45(2):605-11.
11. Rao JN, Scott AJ. A simple method for the analysis of clustered binary data. Biometrics 1992;48(2):577-85.
12. Cochran WG. Ratio estimators. Sampling Techniques. 3rd ed. New York: Wiley; 1977. p.150-86.
13. Heeren T, D'Agostino R. Robustness of the two independent samples t-test when applied to ordinal scaled data. Stat Med 1987;6(1):79-90.
14. Donald A, Donner A. Adjustments to the Mantel-Haenszel chi-square statistic and odds ratio variance estimator when the data are clustered. Stat Med 1987;6(4):491-9.
15. Zhang J, Boos DD. Mantel-Haenszel test statistics for correlated binary data. Biometrics 1997;53(4):1185-98.
16. Liang KY. Odds ratio inference with dependent data. Biometrika 1985;72(3):678-82.
17. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 1959;22(4): 719-48.
18. Song JX, Ahn CW. An evaluation of methods for the stratified analysis of clustered binary data in community intervention trials. Stat Med 2003;22(13):2205-16.
19. Breslow ND, Clayton DG. Approximate inference in generalized linear mixed models. J Am Stat Assoc 1993;88(1):9-25.
20. Wolfinger R, O'Connell M. Generalized linear mixed models: A pseudo-likelihood approach. J Statist Comput Simul 1993;48(3):233-43.
21. Lee Y, Nelder JA. Hierarchical generalized linear models. JR Stat Soc B 1996;58(3):619-78.
22. Begg MD. Analyzing k (2 x 2) tables under cluster sampling. Biometrics 1999;55(1):302-7.
23. Begg MD, Panageas KS. Interval estimation of the common odds ratio from k(2 x 2) tables under cluster sampling. Stat Med 1999;18(9):1087-100.
24. Panageas KS, Begg MD, Grbic JT, Lamster IB. Analysis of multiple 2x2 tables with site-specific periodontal data. J Dent Res 2003;82(7):514-7.
25. Tulunoğlu O, Ulusu T, Işik EE, Tezkirecioğlu M, Genç Y. The prevalence and surface distribution of caries among schoolchildren in Ankara, Turkey according to their dental health behaviors. J Clin Pediatr Dent 2007;31(4):240-5.
26. Ahn C, Jung SH, Donner A. Application of an adjusted chi2 statistic to site-specific data in observational dental studies. J Clin Periodontol 2002;29(1):79-82.