Testing for independence in J × K contingency tables with complex sample survey data

HHS Public Access

Author manuscript

Biometrics. Author manuscript; available in PMC 2016 September 01.

Published in final edited form as:

Biometrics. 2015 September ; 71(3): 832–840. doi:10.1111/biom.12297.

Testing for independence in $J \times K$ contingency tables with complex sample survey data

Stuart R. Lipsitz[1,*], **Garrett M. Fitzmaurice**[2], **Debajyoti Sinha**[3], **Nathanael Hevelone**[1], **Edward Giovannucci**[4], and **Jim C. Hu**[5]

[1]Brigham and Women's Hospital, Boston, MA 02115, USA

[2]Harvard Medical School, Boston, MA 02115, USA

[3]Florida State University, Tallahassee, FL 32306, USA

[4]Harvard School of Public Health, Boston, MA 02115, USA

[5]University of California, Los Angeles, CA 90095, USA

Summary

The test of independence of row and column variables in a ($J \times K$) contingency table is a widely used statistical test in many areas of application. For complex survey samples, use of the standard Pearson chi-squared test is inappropriate due to correlation among units within the same cluster. Rao and Scott (1981) proposed an approach in which the standard Pearson chi-squared statistic is multiplied by a design effect to adjust for the complex survey design. Unfortunately, this test fails to exist when one of the observed cell counts equals zero. Even with the large samples typical of many complex surveys, zero cell counts can occur for rare events, small domains, or contingency tables with a large number of cells. Here, we propose Wald and score test statistics for independence based on weighted least squares estimating equations. In contrast to the Rao-Scott test statistic, the proposed Wald and score test statistics always exist. In simulations, the score test is found to perform best with respect to type I error. The proposed method is motivated by, and applied to, post surgical complications data from the United States' Nationwide Inpatient Sample (NIS) complex survey of hospitals in 2008.

Keywords

Chi-squared test; Nationwide Inpatient Sample; Score statistic; Wald statistic; Weighted estimating equations

[*]slipsitz@partners.org.

7. Supplementary Materials

Web Appendices that include a description of the supplementary materials (Web Appendix A), a SAS macro to calculate the WLS Wald and score statistics as referenced in Section 5 (Web Appendix C), and a data example illustrating the use of the SAS macro (Web Appendix B), are available with this paper at the Biometrics website on the Wiley Online Library.

1. Introduction

A widely used statistical test in many areas of application is the test of independence in a ($J \times K$) contingency table. Even in settings where multivariable multinomial regression analyses are subsequently performed, initial analyses focusing on bivariate associations are regularly reported at the beginning of the results section in published papers. For a typical simple random sample (independent observations), the Pearson chi-squared statistic is widely used and can be shown to equal the score test statistic for testing independence in an $J \times K$ contingency table with row and column variables that are jointly multinomial. It is also equal to the score test statistic for no row (column) covariate effect in a multinomial logistic regression where the column (row) variable is considered the outcome. For complex survey samples, use of the Pearson chi-squared test is not appropriate due to the lack of independence of observations. Many large scale surveys involve stratified, multi-stage sampling and correlation among units within the same cluster.

Recall that complex survey sampling is widely used to sample a fraction of a large finite population while accounting for its size and characteristics. Based on certain subject characteristics (e.g., age, race, gender), some individuals are over or under sampled. Thus, individuals in the population will often have different probabilities of being selected into the sample. Further, the sampling design can be multi-stage. Because of the complex sampling frame utilized in sample surveys, for generalizability of the sample to the finite population (Korn and Graubard, 1999), the design must be incorporated in the analysis, including sampling weights (derived from the probability of selection), strata and/or cluster variables.

A popular test for independence for ($J \times K$) contingency tables with complex survey data has been proposed by Rao and Scott (1981). This approach uses a design effect to adjust the usual Pearson chi-squared statistic for the complex survey design. Unfortunately, this elegant test fails to exist when one of the observed (or weighted) cells in the contingency table equals zero, because the design effect is a function of the inverse of the weighted cell counts. Even with the large samples typical of many complex surveys, zero (weighted) cell counts can occur for rare events, small domains, or contingency tables with a large number of cells, such as in a (5×5) contingency table. With independent observations (without weighting), Fisher's exact test would be the preferred choice with small cell counts. However, Fisher's exact test requires independent observations and cannot be extended to complex survey data due to the stratification, clustering, and weighting.

In this paper, we show that the multinomial logistic regression score test for independence can also be expressed in terms of a linear model for the multinomial outcome; we also show that the score test reduces to a weighted least squares (WLS) estimating equations score test statistic. Further, we describe the properties of the WLS Wald statistic and show that the WLS estimating equations score test statistic can be considered a Wald test statistic with the variance estimated under the null hypothesis.

In Section 2 we introduce some notation for complex sample surveys and discuss polytomous logistic regression and linear regression for the multinomial outcome. In Section 3 we discuss weighted estimating equations (WEE) for complex survey data and present our

proposed WLS Wald and score test statistics. Finally, in Section 4, we present the results of a simulation study examining the finite sample properties of the proposed score and Wald tests in comparison to the Rao-Scott chi-squared test. Before turning to these topics, we introduce an example that motivated the development of the methods.

Our motivating example is from the United States' Healthcare Cost and Utilization Project (HCUP) Nationwide Inpatient Sample (NIS), sponsored by the Agency for Healthcare Research and Quality. With more than 1400 robotic surgical systems installed in hospitals throughout the United States (Yu et al., 2012) robotic-assisted laparoscopic surgery has been rapidly adopted despite the dearth of evidence demonstrating superior outcomes compared to traditional surgical approaches (i.e., non-robotic-assisted laparoscopic surgery and open surgery). The Institute of Medicine has prioritized robotic surgery for comparative effectiveness research (versus the other two types of surgery). There is an absence of population-based studies comparing robotic, laparoscopic and open surgery with respect to surgical complications. In the urological literature (Yu et al., 2012), we recently performed a population-based analyses using the Nationwide Inpatient Sample (NIS) from the last quarter of 2008. Our population-based study objectives are to compare post surgery complications between robotic, laparoscopic and open surgery. In this paper, we focus on nephrectomy (kidney removal) due to kidney cancer.

The NIS is a 20% stratified, cluster probability sample that encompasses approximately 8 million acute hospital stays per year from approximately 1000 hospitals in 37 states. It is the largest all-payer inpatient care observational dataset in the U.S. and is representative of approximately 90% of all hospital discharges. In the NIS, hospitals in the sampling frame are stratified by five key characteristics. Then, a random sample of hospitals (clusters) is chosen from each of the strata. The NIS includes all discharges from the selected hospitals. Each hospital has a different probability of being selected in the sample depending on the five characteristics that determine the strata. As a result, each hospital, and thus all discharges within the hospital, are given a weight so that any results can be extrapolated to the entire universe of hospitals in the United States. Because 20% of the universe of hospitals are sampled, the weights (or inverse probability of being sampled) are usually close to five.

During the last quarter of 2008, there were 3,487 patients with nephrectomy (kidney removal) due to kidney cancer within NIS. The sum of the patient weights equals 12,142; that is, these 3,487 patients represent 12,142 patients in the US population. These 3,487 patients come from 1,051 clusters (hospitals) in 60 strata. For our study, we are primarily interested in determining whether the type of surgery (robotic, laparoscopic, and open) is independent of post-surgery complications (3 levels: no complications, at least one complication without death, death due to surgery). Table 1 presents the 3×3 contingency table of the weighted cell counts (which sum to 12,142); HCUP does not allow publication of observed (unweighted) cell counts between 1 and 12 (although it does allow 0 cell counts) due to concerns about the possibility that patients might be identified. The data in Table 1 suggest that patients who had robotic-assisted laparoscopic surgery have the best complication profile. In Section 5, we present the results of analyses of the NIS data to illustrate the proposed methods.

2. Multinomial Regression with Complex Survey Data

For many complex sample surveys, the target population is usually thought to be of finite size N, and a total of n subjects (or units) are sampled. To indicate which n subjects are sampled from the population of N subjects, we define the indicator random variable $\delta_i = 1$ if subject i is selected into sample, and $\delta_i = 0$ otherwise, for $i = 1, \ldots, N$, where $\sum_{i=1}^{N} \delta_i = n$. Depending on the sampling design, some of the δ_i may be correlated (e.g., for two subjects within the same cluster). We let π_i denote the probability of subject i being selected, which is typically specified in the survey design; π_i may depend on variables of interest, or additional variables (screening variables, for example) not in the analytic model of interest.

Suppose that we are interested in the association between two discrete random variables, X_i and Y_i, where, for subject i in the population, X_i can take on values $1, \ldots, J$, and Y_i can take on values $1, \ldots, K$; to simplify notation, for the remainder of this section we suppress the unit index i. Specifically, we are interested in testing the null hypothesis that X and Y are independent in the population, i.e. H_0: pr[$(X = j)$, $(Y = k)$] = pr$(X = j)$pr$(Y = k)$, for $j = 1, \ldots, J$ and $k = 1, \ldots, K$. Equivalently, we can test for independence between X and Y by treating one of the variables as the outcome and the other as a covariate in a multinomial logistic regression model; the test for no covariate effect is a test for independence (see, for example, Agresti, 2013). Without loss of generality, we treat Y as the outcome and X as the covariate; the test for independence can be rewritten as H_0: pr$(Y = k|X = j)$ = pr$(Y = k)$. We denote these conditional probabilities by $p_{k|j}$ = pr$(Y = k|X = j)$. Note that there are $K - 1$ non-redundant conditional probabilities for each value of $X = j$, since these conditional probabilities must sum to 1 for each level j; so, for simplicity, we discuss the conditional probabilities $p_{k|j}$, $k = 1, \ldots, K - 1$. In terms of these conditional probabilities, the null hypothesis is H_0: $p_{k|j} - p_{k|J} = 0$ for $j = 1, \ldots, J - 1$. This suggests a linear model for the probabilities $p_{k|j}$, with regression coefficients equal to linear contrasts $p_{k|j} - p_{k|j'}$, can be fit and used to test the null hypothesis of independence.

Next, we describe the linear and polytomous logistic regression models for the probabilities $p_{k|j}$. Note that there are $J(K - 1)$ non-redundant conditional probabilities. Any model that has $J(K - 1)$ non-redundant parameters is referred to as a "saturated model". Here, we describe saturated linear and polytomous logistic regression models.

For the saturated linear regression model, we write $p_{k|j}$ as

$$p_{k|j} = p_{k|J} + (p_{k|j} - p_{k|J}) = \alpha_{0k} + \alpha_{jk} \quad (1)$$

for $j = 1, \ldots, J - 1$ and $k = 1, \ldots, K - 1$. The null hypothesis of independence is $\alpha_{jk} = p_{k|j} - p_{k|J} = 0$ for all $j = 1, \ldots, J - 1$ and $k = 1, \ldots, K - 1$. The saturated polytomous logistic regression model for $p_{k|j}$ is

$$\log\left(\frac{p_{k|j}}{p_{K|j}}\right)=\beta_{0k}+\beta_{jk} \quad (2)$$

for $j = 1, \ldots, J-1$ and $k = 1, \ldots, K-1$. The interpretation of the regression parameters is as follows. If we restrict to rows j and J and columns k and K of the contingency table, then β_{jk} is the log odds ratio for the resulting (2×2) table. The null hypothesis of independence is $\beta_{jk} = 0$ for all $j = 1, \ldots, J-1$ and $k = 1, \ldots, K-1$. Transforming to the probability scale,

$$p_{k|j}=\frac{\exp(\beta_{0k}+\beta_{jk})}{1+\sum_{k=1}^{K-1}\exp(\beta_{0k}+\beta_{jk})} \quad (3)$$

for $j = 1, \ldots, J-1$ and $k = 1, \ldots, K-1$.

For both the linear and the polytomous logistic regression model, under the null hypothesis, there are $(J-1)(K-1)$ regression parameters equal to 0; this corresponds to the number of degrees-of-freedom for the standard Pearson chi-squared test of independence for a $(J \times K)$ contingency table. Note, under the null, the intercepts in the linear and polytomous logistic model are related as follows

$$p_{k|j}=\alpha_{0k}=\frac{\exp(\beta_{0k})}{1+\sum_{k=1}^{K-1}\exp(\beta_{0k})}. \quad (4)$$

For estimation and testing, it is convenient to define the indicator random variables, $Y_k = 1$ if $Y = k$ and $Y_k = 0$ otherwise (for $k = 1, \ldots, K$), and $X_j = 1$ if $X = j$ and $X_j = 0$ otherwise (for $j = 1, \ldots, J$). Then, for the linear model,

$$\mathrm{pr}(Y=k|X)=\alpha_{0k}+\sum_{j=1}^{J-1}x_j\alpha_{jk}, \quad (5)$$

and for the polytomous logistic model,

$$\log\left(\frac{p_{k|j}}{p_{K|j}}\right)=\beta_{0k}+\sum_{j=1}^{J-1}x_j\beta_{jk}. \quad (6)$$

Finally, the multinomial probability mass function equals

$$f(y_1,y_2,\ldots,y_K|X)=\prod_{j=1}^{J}\prod_{k=1}^{K}p_{k|j}^{x_jy_k}. \quad (7)$$

In the following section, we propose weighted least squares estimating equations score and Wald tests for independence, naively assuming the multinomial outcomes (the Y_k's) from the same subject are independent in the linear model. Note that the Y_k's from the same

subject are negatively correlated because $\sum_{k=1}^{K} y_k = 1$. Thus, the linear regression score test makes two naive assumptions–that observations within a cluster are independent and the Y_k's from the same subject are independent. In the following section, we also propose a multinomial weighted estimating equations score test for independence based on the polytomous logistic regression model; this test also naively assumes observations within a cluster are independent, but does take into account the multinomial distribution for the Y_k's from the same subject. For both the linear and polytomous logistic score statistics, we use a robust sandwich variance estimator that takes into account that the observations are correlated and consistently estimates the covariance matrix of the score statistics.

3. Weighted Estimating Equations Score Test for Independence

In this section we first discuss the WLS estimating equations for α_{jk} in the linear model, and then the weighted estimating equations (WEE) for β_{jk} in the polytomous logistic regression. Then, we discuss the WLS score and Wald tests for independence from the linear model and the WEE score test for independence from the polytomous logistic regression model.

3.1 Estimating Equations

We let $\boldsymbol{\alpha}_k = (\alpha_{0k}, \alpha_{1k}, \ldots, \alpha_{J-1,k})'$ represent the vector of parameters for the linear model for the k^{th} outcome in (1). To obtain a consistent estimate of $\boldsymbol{\alpha}_k$ with complex survey data, a WLS estimating equations approach can be used, naively assuming the $K - 1$ non-redundant multinomial indicators Y_{ik}, $k = 1, \ldots, K - 1$ are independent, as well as naively assuming that subjects within a cluster are independent. The WEE score vector for outcome y_{ik} is

$$\mathbf{U}(\boldsymbol{\alpha}_k) = \sum_{i=1}^{N} w_i \begin{pmatrix} 1 \\ \mathcal{X}_i \end{pmatrix} \left(y_{ik} - \sum_{j=1}^{J} X_{ij} p_{k|j} \right) = \sum_{i=1}^{N} w_i \begin{pmatrix} 1 \\ \mathcal{X}_i \end{pmatrix} \left(y_{ik} - \alpha_{0k} - \sum_{j=1}^{J-1} x_{ij} \alpha_{jk} \right), \quad (8)$$

$k = 1, \ldots, K - 1$, where $\mathcal{X}_i = (X_{i1}, \ldots, X_{i,J-1})'$ denotes a $(J - 1) \times 1$ vector of the first $J - 1$ non-redundant indicators for X_i, and the complex survey "weights" are $w_i = \frac{\delta_i}{\pi_i}$ ($w_i = \frac{1}{\pi_i}$ if sampled, $\delta_i = 1$). The WLS estimate $\hat{\boldsymbol{\alpha}}_k$ is defined as the simultaneous solution to $\mathbf{U}_k(\hat{\boldsymbol{\alpha}}_k) = 0$, for $k = 1, \ldots, K - 1$. Because $E(y_{ik} - p_{k|j}) = 0$, using method of moments, the WLS estimating equations yield consistent estimators of $\boldsymbol{\alpha}_k$. Even though we have not constrained the $\hat{p}_{k|j}$'s to be between 0 and 1, the non-iterative WLS estimates are

$$\hat{p}_{k|j} = \frac{\sum_{i=1}^{N} w_i x_{ij} y_{ik}}{\sum_{i=1}^{N} w_i x_{ij}}, \quad (9)$$

the weighted proportion of subjects with $Y_i = k$ and $X_i = j$ out of those subjects with $X_i = j$, and $\hat{\boldsymbol{\alpha}}_{jk} = (\hat{p}_{k|j} - \hat{p}_{k|J})$. These estimates are used in a Wald test for H_0: $\alpha_{jk} = 0$ for the linear model. Note, even when there is a zero cell count, i.e., $\hat{p}_{k|j} = 0$, all estimates $\hat{\boldsymbol{\alpha}}_{jk}$ will be finite and a Wald test can be conducted; we discuss this issue later.

Next, we let $\boldsymbol{\beta}_k = (\beta_{0k}, \beta_{1k}, \ldots, \beta_{J-1,k})'$ represent the vector of parameters for the polytomous logistic regression model for the k^{th} outcome in (3). To obtain a consistent estimator of $\boldsymbol{\beta}_k$ with complex survey data, a weighted pseudo-likelihood estimating equations approach can be used; here, the estimating equations take into account the multinomial distribution of the Y_{ik}'s from the same subject. The weighted pseudo-likelihood estimating equations (WEE) score vector for the outcome y_{ik} is

$$\mathbf{S}(\boldsymbol{\beta}_k)=\frac{d}{d\boldsymbol{\beta}_k}\sum_{i=1}^{N}w_i\sum_{j=1}^{J}\sum_{k=1}^{K}x_{ij}y_{ik}\log(p_{k|j})=\sum_{i=1}^{N}w_i\begin{pmatrix}1\\ \mathscr{X}_i\end{pmatrix}\left(y_{ik}-\sum_{j=1}^{J}X_{ij}p_{k|j}\right), \quad (10)$$

for $k = 1, \ldots, K-1$. The pseudo-likelihood score equations have this simple form because the polytomous logistic regression is the canonical model (McCullagh and Nelder, 1989). These estimating equations are identical to the WLS estimating equations in (8), except that $p_{k|j}$ here is expressed in terms of the polytomous logistic regression model. Because the model is saturated (same number of $p_{k|j}$'s as non-redundant model parameters), the estimates of the $p_{k|j}$'s in the saturated model equal those given in (9), and thus are the same as those obtained for the saturated linear model when using WLS. The estimator of β_{jk} is given by,

$$\hat{\beta}_{jk}=\log\left(\frac{\hat{p}_{k|j}}{\hat{p}_{K|j}}\right)-\log\left(\frac{\hat{p}_{k|J}}{\hat{p}_{K|J}}\right).$$

Unfortunately, when there is a zero cell count, i.e., $\hat{p_{k|j}} = 0$, $\hat{\beta_{jk}}$ will equal plus or minus infinity, and a Wald test cannot be performed. Thus, WLS for the linear model and the pseudo-likelihood for the polytomous logistic model yield the exact same form of the estimating equations and identical estimates of $p_{k|j}$ for the saturated model. However, the properties of the Wald statistic will be better when using WLS to estimate the linear model because all estimated parameters will be finite. Thus, for the remainder of the paper, we focus only on the WLS approach.

Similarly, the score statistic, which uses the score vector under the alternative (saturated model) with the $p_{k|j}$'s estimated under the null, is identical using either the WLS for the linear model or the pseudo-likelihood for the polytomous logistic model. For both cases the score vectors are identical, and the estimate of $p_{k|j}$ under the null given in (4) is identical,

$$\tilde{p}_{k|j}=\frac{\sum_{i=1}^{N}w_iy_{ik}}{\sum_{i=1}^{N}w_i}; \quad (11)$$

the latter is simply the weighted proportion of subjects with $Y_i = k$.

3.2 Wald Test

For the linear model, we denote the full $J(K-1)$ vector of parameters as

$$\boldsymbol{\alpha}' = (\boldsymbol{\alpha}_1', \ldots, \boldsymbol{\alpha}_{K-1}').$$

If we denote the $(K-1) \times 1$ vector of multinomial indicators for subject i by $\mathcal{Y}_i = (Y_{i1}, \ldots, Y_{i,K-1})'$, then we can write the linear model for subject i as

$$E(\mathcal{Y}_i | \mathcal{X}_i) = \mathbf{X}_i' \boldsymbol{\alpha},$$

where $\mathbf{X}_i'$ is the $(K-1) \times J(K-1)$ matrix of covariates for subject i,

$$\mathbf{X}_i = I_{K-1} \otimes \begin{pmatrix} 1 \\ \mathcal{X}_i \end{pmatrix},$$

$\mathbf{I}_{K-1}$ is a $K-1$ identity matrix, and $\otimes$ is the Kronecker product symbol. Then, the WLS estimate is non-iterative:

$$\hat{\boldsymbol{\alpha}} = \left(\sum_{i=1}^{N} w_i \mathbf{X}_i \mathbf{X}_i' \right)^{-1} \sum_{i=1}^{N} w_i \mathbf{X}_i \mathcal{Y}_i.$$

For the WLS approach for the linear model, we denote the full score vector as

$$\mathbf{U}(\boldsymbol{\alpha}) = \{ \mathbf{U}(\boldsymbol{\alpha}_1)', \ldots, \mathbf{U}(\boldsymbol{\alpha}_{K-1})' \}' = \sum_{i=1}^{N} w_i \mathbf{X}_i (\mathcal{Y}_i - \mathbf{X}_i' \boldsymbol{\alpha}).$$

Using a first-order Taylor series expansion and a suitable central limit theorem for sample survey data (Binder, 1983), the WLS estimate $\hat{\boldsymbol{\alpha}}$ is consistent and has an asymptotic multivariate normal distribution with mean $\boldsymbol{\alpha}$ and covariance matrix

$$\left\{ E\left(\frac{d\mathbf{U}(\boldsymbol{\alpha})}{d\boldsymbol{\alpha}} \right) \right\}^{-1} Var\{\mathbf{U}(\boldsymbol{\alpha})\} \left\{ E\left(\frac{d\mathbf{U}(\boldsymbol{\alpha})}{d\boldsymbol{\alpha}} \right) \right\}^{-1}, \quad (12)$$

where,

$$E\left(\frac{d\mathbf{U}(\boldsymbol{\alpha})}{d\boldsymbol{\alpha}} \right) = H = \sum_{i=1}^{N} w_i \mathbf{X}_i \mathbf{X}_i', \quad (13)$$

Note, although (13) holds for any sampling design, the form of $Var\{\mathbf{U}(\boldsymbol{\alpha})\}$ in (12) depends on the sampling scheme.

For an example of how to estimate $Var\{\mathbf{U}(\boldsymbol{\alpha})\}$ for a given sampling scheme, suppose we have a design with independent clusters sampled from S strata, and for simplicity, suppose the stratum population sizes are sufficiently large that no finite sample correction is required. With a slight change in notation, suppose there are $s = 1, \ldots, S$ strata, $\ell = 1, \ldots, n_s$ clusters within stratum s, $m = 1, \ldots, M_{s\ell}$ subjects within cluster ℓ of stratum s, and a total of $n=\sum_{s=1}^{S}\sum_{\ell=1}^{n_s} M_{s\ell}=\sum_{i=1}^{N}\delta_i$ subjects in the sample. Then, a consistent estimate of $Var\{\mathbf{U}(\boldsymbol{\alpha})\}$ in (12) is

$$\mathbf{G}=\frac{n-1}{n-J(K-1)}\sum_{s=1}^{S}\frac{n_s}{n_s-1}\sum_{\ell=1}^{n_s}\{\mathbf{U}_{s\ell+}(\hat{\boldsymbol{\alpha}})-\overline{\mathbf{U}}_{s++}(\hat{\boldsymbol{\alpha}})\}\{\mathbf{U}_{s\ell+}(\hat{\boldsymbol{\alpha}})-\overline{\mathbf{U}}_{s++}(\hat{\boldsymbol{\alpha}})\}', \quad (14)$$

where

$$\mathbf{U}_{s\ell m}(\hat{\boldsymbol{\alpha}})$$

is the contribution to the score vector from the m^{th} subject within cluster ℓ of stratum s,

$$\mathbf{U}_{s\ell+}(\hat{\boldsymbol{\alpha}})=\sum_{m=1}^{M_{s\ell}}\mathbf{U}_{s\ell m}(\hat{\boldsymbol{\alpha}})$$

is the sum of the score vectors from the sample subjects in cluster ℓ of stratum s and

$$\overline{\mathbf{U}}_{s++}(\hat{\boldsymbol{\alpha}})=\frac{1}{n_s}\sum_{\ell=1}^{n_s}\mathbf{U}_{s\ell+}(\hat{\boldsymbol{\alpha}})$$

is the mean of the $\mathbf{U}_{s\ell+}(\hat{\boldsymbol{\alpha}})$'s. Further, using the results of Morel (1989), an approximately unbiased estimate of $Var\{\mathbf{U}(\hat{\boldsymbol{\alpha}})\}$ in small samples is

$$\hat{\mathrm{Var}}\{\mathbf{U}(\hat{\boldsymbol{\alpha}})\}=\mathbf{G}+\gamma\phi\mathbf{H}, \quad (15)$$

where

$$\gamma=\max\left\{1,\mathrm{trace}\left(\mathbf{H}^{-1}\mathbf{G}\right)\right\}$$

and

$$\phi=\min[0.5,\{J(K-1)\}/\{n-J(K-1)\}].$$

Thus, an approximately unbiased, consistent estimate (Morel, 1989) of the variance of $\hat{\boldsymbol{\alpha}}$ is

$$\hat{\mathrm{Var}}(\hat{\boldsymbol{\alpha}})=\mathbf{H}^{-1}(\mathbf{G}+\gamma\phi\mathbf{H})\mathbf{H}^{-1}. \quad (16)$$

The term $\gamma\phi\mathbf{H}$ in (15) guarantees that (16) is positive definite as long as **H** is invertible; without this extra term, it is possible that (16) is positive semi-definite.

For the vector $\boldsymbol{\alpha}_k$, the null hypothesis of independence for the linear model H_0: $\alpha_{jk} = p_{k|j} - p_{k|J} = 0$, for $j = 1, \ldots, J-1$ is a linear contrast of the form

$$H_0:\mathbf{r}\boldsymbol{\alpha}_k=0$$

where the $J-1 \times J$ matrix **r** equals

$$\mathbf{r}=\begin{pmatrix} \mathbf{0} & \mathbf{I}_{J-1} \end{pmatrix},$$

and $\mathbf{I}_{J-1}$ is a $J-1 \times J-1$ identity matrix and **0** is a $J-1$ vector of zeros. For the full vector $\boldsymbol{\alpha}$, the null hypothesis is a linear contrast of the form

$$H_0:\mathbf{R}\boldsymbol{\alpha}=0$$

where the $(J-1)(K-1) \times J(K-1)$ matrix $\mathbf{R} = \mathbf{I}_{K-1} \otimes \mathbf{r}$.

For a stratified, cluster design with S strata and C clusters, the general Wald statistic proposed by Korn and Graubard (1999) for testing $H_0:\mathbf{R}\boldsymbol{\alpha} = 0$ is

$$F_W=\frac{(C-S)-(J-1)(K-1)+1}{(C-S)(J-1)(K-1)}(\mathbf{R}\hat{\boldsymbol{\alpha}})'\{\mathbf{R}\hat{\mathrm{Var}}(\hat{\boldsymbol{\alpha}})\mathbf{R}'\}^{-1}\mathbf{R}\hat{\boldsymbol{\alpha}}, \quad (17)$$

which has an F-distribution with $(J-1)(K-1)$ and $(C-S)-(J-1)(K-1)+1$ degrees-of-freedom under the null. We use this latter approximation due to the "small sample" data configurations for which we are proposing use of our test statistic.

3.3 Score Test

To develop the WLS score statistic, we rewrite the WLS score vector for the linear model in matrix terms as

$$\mathbf{U}(\boldsymbol{\alpha})=\mathbf{X}'\mathbf{W}(\mathbf{y}-\mathbf{X}\boldsymbol{\alpha}) \quad (18)$$

where **X** is a $n(K-1) \times n(K-1)$ matrix containing the covariates for all $(K-1)$ outcomes for all subjects in the study, **W** is an $n(K-1) \times n(K-1)$ matrix containing the weights for all subjects, and **y** is an $n(K-1) \times 1$ vector containing all $(K-1)$ outcomes for all subjects in the study. Note, we can rewrite (18) as

$$\mathbf{U}(\boldsymbol{\alpha})=(\mathbf{X}'\mathbf{W}\mathbf{X})(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}), \quad (19)$$

where $\hat{\boldsymbol{\alpha}} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}$ is the WLS estimate from the full model.

The WLS score statistic is based on the large sample distribution of

$$\mathbf{U}(\tilde{\boldsymbol{\alpha}})=(\mathbf{X}'\mathbf{W}\mathbf{X})(\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}}), \quad (20)$$

where $\tilde{\boldsymbol{\alpha}}$ is the constrained estimate of $\boldsymbol{\alpha}$ under the null hypothesis that $\mathbf{R}\boldsymbol{\alpha} = 0$. To develop the WLS score test, we use the Lagrange multiplier form of the score statistic, since the score test and Lagrange multiplier test have been shown to be equivalent (Bera and Bilias, 2001). The Lagrangian function (Silvey, 1959) is defined as

$$\mathscr{L}(\boldsymbol{\alpha},\boldsymbol{\lambda})=\mathbf{U}(\boldsymbol{\alpha})-\mathbf{R}'\boldsymbol{\lambda}=(\mathbf{X}'\mathbf{W}\mathbf{X})(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha})-\mathbf{R}'\boldsymbol{\lambda} \quad (21)$$

for the $(J-1)(K-1)\times 1$ vector of Lagrange multipliers $\boldsymbol{\lambda}$. Subject to the constraint $\mathbf{R}\tilde{\boldsymbol{\alpha}} = 0$, the estimate of $\boldsymbol{\alpha}$ under the null, $\tilde{\boldsymbol{\alpha}}$, can be obtained as the solution to

$$\mathscr{L}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\lambda}})=(\mathbf{X}'\mathbf{W}\mathbf{X})(\hat{\boldsymbol{\alpha}}-\tilde{\boldsymbol{\alpha}})-\mathbf{R}'\hat{\boldsymbol{\lambda}}=0. \quad (22)$$

The Lagrange multiplier test for H_0:$\mathbf{R}\boldsymbol{\alpha} = 0$ can be written as H_0: $\boldsymbol{\lambda} = 0$.

Using the results of Amemiya (1985), the constrained WLS estimate $\tilde{\boldsymbol{\alpha}}$ equals

$$\tilde{\boldsymbol{\alpha}}=\hat{\boldsymbol{\alpha}}-(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{R}'\{\mathbf{R}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{R}'\}^{-1}\mathbf{R}\hat{\boldsymbol{\alpha}}. \quad (23)$$

Plugging (23) in (22), we obtain

$$\mathbf{R}'\{\mathbf{R}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{R}'\}^{-1}\mathbf{R}\hat{\boldsymbol{\alpha}}-\mathbf{R}'\hat{\boldsymbol{\lambda}}=0, \quad (24)$$

from which it follows that

$$\hat{\boldsymbol{\lambda}}=\{\mathbf{R}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{R}'\}^{-1}\mathbf{R}\hat{\boldsymbol{\alpha}}$$

The general form of the Lagrange multiplier (score) statistic for testing H_0:$\mathbf{R}\boldsymbol{\alpha} = 0$ is

$$\hat{\boldsymbol{\lambda}}'\widehat{\mathrm{Var}}(\hat{\boldsymbol{\lambda}})^{-1}\hat{\boldsymbol{\lambda}}. \quad (25)$$

However, since

$$Var(\hat{\boldsymbol{\lambda}})=\{\mathbf{R}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{R}'\}^{-1}\mathbf{R}\,Var(\hat{\boldsymbol{\alpha}})\mathbf{R}'\{\mathbf{R}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{R}'\}^{-1}, \quad (26)$$

when substituting (26) in (25), it is easy to show that (25) reduces to

$$(\mathbf{R}\hat{\boldsymbol{\alpha}})'\{\mathbf{R}\hat{\mathrm{V}}\mathrm{ar}(\hat{\boldsymbol{\alpha}}|H_0)\mathbf{R}'\}^{-1}\mathbf{R}\hat{\boldsymbol{\alpha}}. \quad (27)$$

Typically, in a Lagrange multiplier (score) statistic, $Var(\hat{\boldsymbol{\alpha}})$ is estimated under the null hypothesis, which we denote by $\hat{\mathrm{V}}\mathrm{ar}(\hat{\boldsymbol{\alpha}}|H_0)$ in (27). In general, for a stratified, cluster design with *S* strata and *C* clusters, the score statistic is

$$F_W=\frac{(C-S)-(J-1)(K-1)+1}{(C-S)(J-1)(K-1)}(\mathbf{R}\hat{\boldsymbol{\alpha}})'\{\mathbf{R}\hat{\mathrm{V}}\mathrm{ar}(\hat{\boldsymbol{\alpha}}|H_0)\mathbf{R}'\}^{-1}\mathbf{R}\hat{\boldsymbol{\alpha}}, \quad (28)$$

which has an *F*-distribution with $(J-1)(K-1)$ and $(C-S)-(J-1)(K-1)+1$ degrees-of-freedom under the null (Rao, Scott, and Skinner, 1998). The estimate $\hat{\mathrm{V}}\mathrm{ar}(\hat{\boldsymbol{\alpha}}|H_0)$ is obtained by replacing $\hat{\boldsymbol{\alpha}}$ in (14) with $\tilde{\boldsymbol{\alpha}}$. Thus, we see that both the Wald statistic in (17) and the score statistic in (28) are quadratic forms in the observed contrast $\mathbf{R}\hat{\boldsymbol{\alpha}}$, so that we might expect them to perform similarly. In fact, the score statistic can be considered a Wald statistic with the variance estimated under the null.

For non-linear regression models such as logistic regression, the Wald test is known to exhibit unreliable and aberrant behavior (for example, with zero cell counts the estimated parameters are not finite); as a result, the score test is preferred for logistic regression in smaller samples. For WLS with a linear model, the parameters are always finite, and the Wald and score statistics are similar quadratic forms in $\mathbf{R}\hat{\boldsymbol{\alpha}}$, except that the score statistic has the variance estimated under the null. Thus, any differences in performance of the WLS Wald and score statistics will be dependent on how well the variance is estimated for each. This is explored further in the following section where we perform a simulation study of the finite sample properties of the test statistics.

4. Simulation Study

We conducted a simulation study primarily to explore the finite sample properties of the proposed WLS Wald and score test statistics for the linear model versus the Rao-Scott chi-squared test. For simplicity, in the simulation study, we used a cluster design without stratification where individuals within a cluster had different probabilities of selection, and thus different weights.

Specifically, we considered a (4 × 3) contingency table, where the row variable $X = 1, 2, 3, 4$ is a cluster-level variable, and given *X* the column variable $Y = 1, 2, 3$ for a subject within a cluster follows a Dirichlet-multinomial distribution (Mosimann, 1962). In the Dirichlet-multinomial distribution, the multinomial column probabilities follow a Dirichlet distribution, and given these probabilities and *X*, the distribution of *Y* follows a multinomial distribution. Also, in the Dirichlet-multinomial distribution, the intracluster correlation for

two subjects in the same cluster is denoted by ρ, and we specify the marginal probability that $Y = k$ given $X = j$ using an adjacent category polytomous logistic regression model (Haberman, 1974). In general, the adjacent categories logit is a special case of the polytomous logistic regression model discuss earlier. Using the last level $Y = K$ as reference, the j^{th} adjacent category logit is

$$\log\left(\frac{p_{k|j}}{p_{K|j}}\right)=\beta_{0k}+\beta(K-k)(J-j),$$

for $j = 1, \ldots, J - 1$ and $k = 1, \ldots, K - 1$. Although the adjacent category logit is a model for ordinal data, we specified the model in this manner to simplify the description of the alternative with a single association parameter (β). This specification allows us to easily explore power as a function of β.

For a given simulation configuration, we generated 2000 clusters from the Dirichlet-multinomial distribution, and considered this to be the 'finite population'. In the population, 25% of the clusters have each of the values of X; however, based on the selection probabilities below, the proportion of subjects with each value of X in a given simulation sample could be different than 25%. For a given configuration, we fixed the population cluster size at $M = 10$ and the population intracluster correlation at $\rho = 0.02$. We note that the marginal probabilities from the adjacent category logit and the population intracluster correlation completely specify the Dirichlet-multinomial distribution.

From this finite population, we sampled clusters with different probabilities based on the value of X for that cluster. Then, once a cluster was sampled, individuals within each cluster were sampled with selection probabilities that varied as a function of both Y and X. This type of sampling design is commonly used in complex surveys where, for example, the cluster is the household, and X is a cluster level variable that represents quartiles of household income and Y represents a variable that is individual specific, such as gender. If we denote the selection probability by $\pi_{jk} = pr(\delta = 1|X = j, Y = k)$, then for all simulations,

$$(\pi_{11},\pi_{12},\pi_{13},\pi_{21},\pi_{22},\pi_{23},\pi_{31},\pi_{32},\pi_{33},\pi_{41},\pi_{42},\pi_{43})=(0.030,0.023,0.015,0.038,0.031,0.025,0.03,0.026,0.022,0.025,0.022,0.0$$

Thus, although the cluster sizes $M = 10$ were the same in the population, the cluster sizes varied in the sample because π_{jk} varied as a function of Y and X.

We explored both the type I error and the power of the WLS score test statistic, the Rao-Scott chi-squared statistic, and the WLS Wald test statistic. Table 2 presents the rejection percentages as well as the percent of simulations in which the test statistic exists (recall, if a cell count is zero, the Rao-Scott chi-squared statistic does not exist) for two simulation sets in which at least one $p_{k|j} \leq 0.02$ under the null. For the first simulation set at the top of Table 2, we let ($\beta_{01} = -4$, $\beta_{02} = -0.5$) so that, under the null, $Y = 1$ has a small probability: ($p_{1|j}$, $p_{2|j}$, $p_{3|j}$) = (0.011, 0.37, 0.62) for $j = 1, \ldots, 4$. We set $\beta = 0$ to examine the type 1 error and then increased β to study power. As can be seen at the top of Table 2, the type I error for the

WLS score statistic is approximately 5%, and the test statistic always existed. The WLS Wald statistic also always existed, but had relatively high type I error. The type I error for the Rao-Scott chi-squared statistic is approximately 5% in the simulations for which it existed, but we also see that it only existed in approximately 27% of the simulation replications. Comparing the score and Wald statistics with respect to power, we see that the Wald statistic had the highest power, but that was to be expected given that its type I error is high. We see that as β increases, the Rao-Scott statistic exists for a higher percentage of simulation replications, but still only exists at most 78% of the time for any given value of β.

Table 2 (bottom) presents the rejection percentages when ($\beta_{01} = -4$, $\beta_{02} = -4$) so that, under the null, both $Y = 1$ and $Y = 2$ have small probabilities: ($p_{1|j}$, $p_{2|j}$, $p_{3|j}$) = (0.02, 0.02, 0.96) for $j = 1, \ldots, 4$. The type I error for the Rao-Scott chi-squared statistic is very low, and we also see that it existed in approximately 50% of the simulation replications (this percentage is higher than under the null for ($p_{1|j}$, $p_{2|j}$, $p_{3|j}$) = (0.011, 0.37, 0.62) which has only one small probability; we conjecture this is because $p_{1|j}$ and $p_{2|j}$ are both larger than $p_{1|j}$ in the first simulation configuration). Again, the Wald statistic has high type I error and the highest power, whereas the score statistic has type I error close to the nominal 5% value. Also, as β increases, the Rao-Scott statistic exists for a higher percentage of simulation replications, but still only exists at most 71% of the time for any given value of β.

For less rare events, Table 3 presents the rejection percentages when ($\beta_{01} = 0$, $\beta_{02} = 0$) so that, under the null, the 3 possible values of Y have equal probability: ($p_{1|j}$, $p_{2|j}$, $p_{3|j}$) = (0.33, 0.33, 0.33) for $j = 1, \ldots, 4$. The test statistics exist for all approaches (including the Rao-Scott statistic) for all simulations replications, so these percentages are not reported in the table. In Table 3, we see that the type I error for the Rao-Scott and Wald statistics are slightly high, whereas the score statistic has type I error close to the nominal 5% value.

Overall, the results of this simulation study suggests that the WLS score statistic may be preferred, since it has good properties relative to the Rao-Scott statistic in the simulations with non-rare events where the Rao-Scott statistic exists, and has discernibly better properties (always exists and has the correct type I error rate) in simulations with rare events in which the Rao-Scott statistic does not exist. Also, the type I error rate for the WLS Wald statistic appears high in all simulation configurations displayed.

5. Application: NIS Nephrectomy Study

Next, we present results of analyses of data from the NIS nephrectomy example discussed in the Introduction. Although the sampling of hospitals (clusters) was performed without replacement in each stratum, the total (population) number of clusters within each stratum was sufficiently large that the finite population correction factor can safely be ignored. The weights used in the analysis are the (Horvitz-Thompson) survey weights provided by NIS, so that the weights sum to the population total. These weights also account for unit non-response. Our goal is to test for independence between the type of surgery (robotic, laparoscopic, open) and post-surgery complications (3 levels: no complications, at least one complication without death, death due to surgery). Table 1 presents the 3×3 contingency table of weighted cell counts. Examination of Table 1 reveals that no patient who received

robotic surgery had a post-surgical death; further, robotic surgery appears to have a better complication profile than the other two types of surgery. Because of the zero cell count, the Rao-Scott chi-squared statistic is not computable. Instead, we computed the WLS score and Wald statistics proposed in Section 3; a SAS macro to calculate the WLS Wald and score statistics is included as supplementary material for this paper at the Biometrics website.

The WLS estimating equation score statistic had an F-value (with 4 and 991 degrees-of-freedom) of 5.22, with *P*–value < 0.001. The Wald test statistic had an F = 5.96, $P < 0.001$. Here both statistics indicate that there is strong evidence that type of surgery is related to post-surgery complications. When each of the three 2 degrees-of-freedom contrasts for the pairwise comparisons of surgery type are considered, the dependence is primarily due to differences in rates of post-surgery complications with open surgery versus each of the other two types of surgery. We found significant differences when comparing open surgery to robotic-assisted laparoscopic surgery (score statistic F(2, 991) = 7.34, $P < 0.001$; Wald statistic F(2, 991) = 9.94, $P < 0.001$); as well as when comparing open surgery to non-robotic-assisted laparoscopic surgery (score statistic F(2, 991) = 5.42, $P < 0.005$; Wald statistic F(2, 991) = 5.69,$P < 0.005$). There were no significant differences when comparing robotic-assisted to non-robotic-assisted laparoscopic surgery (score statistic F(2, 991) = 2.23, $P = 0.108$; Wald statistic F(2, 991) = 2.41, $P = 0.090$). In summary, results of the analyses indicate that there are significantly more post-surgery complications with open surgery.

6. Conclusion

In this paper we propose weighted least squares score and Wald tests for independence with complex survey data. The proposed approach is not *ad hoc*, but is based on theory for estimating equations score and Wald test statistics (Rao et al., 1998). Results of our simulation study suggest that the proposed score test statistic has better properties than the Rao-Scott test statistic in that the score test statistic always exists, and appears to have the correct type I error. In addition, the results of the simulations suggest that the WLS Wald test statistic has high type I error. In addition to having better type I error properties, the score test may also be preferred due to the fact that the Wald test is not invariant under reparameterization. We note that in additional simulations (results not reported), for larger cluster sizes, the Wald test statistic has the correct type I error. Finally, we note that the approach used in this paper to develop a score test could also be used to formulate a test of independence in other settings, e.g., an extension of the Cochran-Mantel-Haenszel statistic for conditional independence of stratified categorical data to the complex survey setting.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful for the support provided by grants MH 054693, CA 160679 and CA 06922 from the U.S. National Institutes of Health.

References

Agresti, A. Categorical Data Analysis. 3. New York: Wiley; 2013.

Amemiya, T. Advanced Econometrics. Harvard University Press; 1985.

Aitchison J, Silvey SD. Maximum-likelihood estimation of parameters subject to restraints. Ann Math Stat. 1958; 29:813–828.

Bera AK, Bilias Y. Raos score, Neymans C(α) and Silveys LM tests: An essay on historical developments and some new results. Journal of Statistical Planning and Inference. 2001; 97:9–44.

Boos DD. On generalized score tests. The American Statistician. 1992; 46:327–333.

Goodman LA. The analysis of cross-classified data having ordered and/or unordered categories: Association models, correlation models, and asymmetry models for contingency tables with or without missing entries. Annals of Statistics. 1985; 13:10–69.

Hauck WW, Donner A. Wald's test as applied to hypotheses in logit analysis. Journal of the American Statistical Association. 1977; 72:851–853.

Korn, EL.; Graubard, BI. Analysis of Health Surveys. New York: Wiley; 1999.

McCullagh, P.; Nelder, JA. Generalized Linear Models. 2. New York: Chapman and Hall; 1989.

Morel JG. Logistic regression under complex survey designs. Survey Methodology. 1989; 15:203–223.

Mosimann JE. On the compound multinomial distribution, the multivariate *beta*-distribution, and correlation among proportions. Biometrika. 1962; 49:65–82.

Rao JNK, Scott AJ. The analysis of categorical data from complex surveys: chi-squared tests for goodness of fit and independence in two-way tables. Journal of the American Statistical Association. 1981; 76:221–230.

Rao JNK, Scott AJ, Skinner CJ. Quasi-score tests with survey data. Statistica Sinica. 1998; 8:1059–1070.

Rotnitzky A, Jewell NP. Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. Biometrika. 1990; 77:485–497.

Silvey SD. The Lagrangian multiplier test. Ann Math Stat. 1959; 30:389–407.

Yu HY, Hevelone ND, Lipsitz SR, Kowalczyk KJ, Nguyen PL, Choueiri TK, Kibel AS, Hu JC. Comparative analysis of outcomes and costs following open radical cystectomy versus robot-assisted laparoscopic radical cystectomy: results from the US Nationwide Inpatient Sample. Eur Urol. 2012; 61:1239–1244. [PubMed: 22482778]

Table 1

Surgery type by complication status for NIS nephrectomy data, weighted counts and row percentages.

	Complication Status		
Surgery Type	None	≥ 1	Death
Robotic	166	71	0
	70.0%	30.0%	0.0%
Laparoscopic	1476	742	13
	66.2%	33.2%	0.6%
Open	6059	3476	139
	62.6%	35.9%	1.5%

Table 2

Simulations results for weighted, clustered data, with $(p_{1|j}, p_{2|j}, p_{3|j}) = (0.011, 0.37, 0.62)$ and $(p_{1|j}, p_{2|j}, p_{3|j}) = (0.02, 0.02, 0.96)$ under the null.

		Rao-Scott		WLS Score		WLS Wald	
$(p_{1\|j}, p_{2\|j}, p_{3\|j})$	β	% Rejected	% Exist	% Rejected	% Exist	% Rejected	% Exist
(0.011,0.37,0.62)	0.00	4.8	27.3	5.8	100	9.1	100
	0.10	15.6	46.6	11.3	100	17.6	100
	0.125	26.1	47.6	29.1	100	38.9	100
	0.15	43.1	63.6	42.4	100	53.4	100
	0.20	68.2	69.3	72.2	100	78.8	100
	0.25	89.5	78.4	92.7	100	95.9	100
(0.02,0.02,0.96)	0.00	1.0	50.2	4.1	100	11.3	100
	0.10	3.0	63.2	8.1	100	18.3	100
	0.20	12.7	65.2	17.8	100	29.3	100
	0.25	32.3	71.0	32.0	100	43.8	100
	0.30	69.9	70.0	62.2	100	72.1	100
	0.35	83.7	67.1	80.7	100	87.5	100

Table 3

Simulations results for weighted, clustered data, with $(p_{1|j}, p_{2|j}, p_{3|j}) = (0.33, 0.33, 0.33)$ under the null.

	Rao-Scott	WLS Score	WLS Wald
β	% Rejected	% Rejected	% Rejected
0.00	6.8	4.3	6.8
0.05	14.8	10.1	15.7
0.08	27.4	19.9	25.2
0.10	40.8	32.9	44.6
0.15	65.7	62.1	72.2
0.20	92.7	92.8	95.6

Note: Test statistics exist for all approaches for all simulations