APPLICATION OF LOGISTIC MODELS TO SURVEY DATA USING REPLICATION TECHNIQUES

Leyla Mohadjer, Ibrahim Yansaneh, J. Michael Brick, Westat Inc.
Leyla Mohadjer, Westat Inc., 1650 Research Boulevard, Rockville, Maryland 20850

1. Introduction

The logistic regression model is being used in many different areas and has become the standard method of analyzing models in which the dependency of a binary response variable is being tested on a number of explanatory variables. In practice, most of the data used in these analyses are obtained from surveys with complex designs, in which stratification, clustering and multi-stages of sampling are used to collect the data. There is an extensive body of literature on the effect of complex survey designs on the analysis of such data. The literature includes various methods of accounting for the sample design when analyzing the data. However, the literature on analyzing logistic regression models using survey data is limited. This paper provides a discussion on the use of replication (resampling) techniques in analyzing logistic models including the application of replication techniques to the score test method. Westat, Inc. has recently enhanced the WesVarPC software by adding analysis of logistic regression models with survey data using replication techniques. A description of logistic regression modeling in WesVarPC is provided.

In simple random sampling, a logistic model is defined in the following way. Suppose that a response (dependent) variable Y can take one of the two values 0 or 1, i.e., occurrence or nonoccurrence of an event such as injury in an accident. Variables of this type are often called binary or dichotomous variables. For dichotomous variables such as Y, one object is to develop a method for estimating P, where P is the probability of occurrence of an event as a function of a number of independent variables. It has been shown, theoretically and empirically, that when the dependent variable is dichotomous, the shape of the response function is frequently curvilinear. The logistic regression model is a curvilinear response function which has been found to be appropriate in many cases involving a binary dependent variable. This response function assures that the estimate of P (probability of occurrence of an event) is always between 0 and 1.

Let Y_i denote the observed value of Y for the i-th individual in the sample, and designate the corresponding vector of p regressor (independent) variables X_i, $i = 1, 2,..., n$, where

$$\mathbf{X}_i' = (X_{0i}, X_{1i}, \ldots, X_{pi}).$$

The dummy regressor $X_{0i} = 1$ $(i = 1,..., n)$ is included to provide for the estimation of an intercept. Also define the n x $(p + 1)$ matrix of independent variables for the sample as X, where

$$\mathbf{X}' = (\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n).$$

Then, under the logistic regression model, the probability that Y_i is equal to 1 is

$$P_i = 1/(1 + \exp(-\beta' \mathbf{X}_i)), \tag{1}$$

where $\beta' = \left(\beta_0, \beta_1, ..., \beta_p\right)$ is the vector of regression coefficients in the logistic model. Equation (1), can be written in the following way

$$Y_i = 1/(1 + \exp(-\beta' \mathbf{X}_i)) + \varepsilon_i, \tag{2}$$

where ε_i is a random error with mean zero and variance P_i $(1 - P_i)$. Refer to Kleinbaum, Kupper and Morgenstern (1982), pages 421-446, and Hosmer and Lemeshow (1989) for more information about logistic regression models.

Section 2 provides a brief overview of applying logistic models to survey data. A discussion on the use of replication (resampling) techniques in analyzing logistic models is given in Section 3. A brief description of the analysis of logistic models in WesVarPC is given in Section 4. A summary of an empirical study on the application of replication techniques for the computation of score test is given in Section 5.

2. Applying Logistic Models to Survey Data

The use of survey data to analyze regression models, especially of the association between risk factors and covariates such as sets of characteristics and outcomes, has become increasingly more popular among data analysts. In these cases, a model is postulated, the parameters of the model are estimated, and tests of hypotheses of the fit of the model or a subset of the model parameters are conducted. To carry on such analyses, it is necessary to measure the precision of the estimated parameters. The sample design used to select survey data are usually complex, involving multistage sampling, clustering, and differential sampling probabilities for various subgroups in the sample. There is a body of literature on whether the sample weights and sample design should be accounted for in model fitting. In general, regression modeling with survey data is accomplished by either using a design-based or a model-based approach. The material provided in this paper is solely related to the design-based approach in which the effect of the sample design on the covariance structure of the sample data is taken into account when fitting logistic models. The reader is referred to Pfeffermann (1993) for a survey of relevant literature focusing on the design-based approach. The book entitled "Model Assisted Survey Sampling" by Sarndal, Swenson, Wretman, (1992) provides a general reference for the model-based approach.

The Pseudo likelihood approach to the analysis of logistic models utilizes the sampling weights to estimate the likelihood equations that would have been obtained in the case of a census. Let p = number of independent variables, and n = sample size. We define $\mathbf{W}$ to be the n x n diagonal matrix formed from the elements of the $\mathbf{w}$, where $\mathbf{w}$ is the n x 1 vector of the full sample weights $\left[w_1, w_2, \ldots, w_n\right]$ associated with the n observations in the sample.

Following the notation in equation (2), the "weighted" maximum likelihood estimates of β' denoted by $\mathbf{b}'$ are the solutions to the equations

$$\mathbf{X}' \, \mathbf{W} \, (\mathbf{Y} - \hat{\mathbf{P}}) = 0 \tag{3}$$

where $\mathbf{Y}$ denotes the vector of observed values of the response variable

$$\mathbf{Y}' = (Y_1, Y_2, \ldots, Y_n),$$

and $\hat{\mathbf{P}} = \left(\hat{P}_1, \hat{P}_2, \ldots, \hat{P}n\right)$ is the associated vector of estimated probabilities.

The iterative method of Newton-Raphson is used to solve for

$$\mathbf{b}' = \left(b_0, b_1, b_2, \ldots, b_p\right).$$

The initial parameter estimates are taken to be zeros unless starting values are specified. The convergence is obtained when the difference in -2 log-likelihood between successive steps is less than some prespecified value, where the weighted estimate of the log-likelihood is given by

$$\sum_{i=1}^{n} w_i Y_i \log \hat{P}_i + \sum_{i=1}^{n} w_i \left(1 - Y_i\right) \log\left(1 - \hat{P}_i\right).$$

3. Replication Techniques for Logistic Models

As noted earlier, it is necessary to measure the precision of the parameters estimated for a model when conducting tests of hypotheses for the fit of the model. Two common methods are available for estimation of variances when data comes from a complex sample design. One is linearization in which nonlinear estimates are approximated by linear ones for the purpose of variance estimation. The linear approximation is derived by taking the first order Taylor series approximation for the estimator. The second method is replication, in which several estimates of the population parameter of interest are derived from a number of subsamples of the original sample. The variability of these replicated estimates is used to estimate the variance of the estimator. See Wolter (1985) for a description of both of these approaches. This paper uses the replication approach.

For logistic models, a replication approach computes "weighted" maximum-likelihood estimates of the parameters and applies the balanced repeated replication method (McCarthy 1969) or jackknife method (Wolter 1985) to estimate the sampling errors of the model parameters.

Denote the replicate estimates of β by $\mathbf{b}_{(k)}$, $k = 1,\ 2\ldots,\ K$. Then an estimate of the variance-covariance matrix of $\mathbf{b}$ based on replicates is given by

$$\hat{\mathrm{Var}}(\mathbf{b}) = c \sum_{k} (\mathbf{b}_{(k)} - \mathbf{b})\,(\mathbf{b}_{(k)} - \mathbf{b})'$$

where c is a constant that depends on the replication approach.

Assuming that the vector of estimated coefficients, $\mathbf{b}$, has approximately a multivariate normal sampling distribution with mean β and variance-covariance matrix $\mathrm{Var}(\mathbf{b})$, the hypothesis

$$H_0\colon\ \mathbf{D}\beta = \delta \text{ versus}$$
$$H_1\colon\ \mathbf{D}\beta \neq \delta$$

can be tested using the statistic

$$T_d^2 = (\mathbf{Db} - \delta)' \, (\mathbf{D} \, \hat{\mathrm{Var}}(\mathbf{b}) \, \mathbf{D}')^{-1} \, (\mathbf{Db} - \delta).$$

If the degrees of freedom associated with $\hat{\mathrm{Var}}(\mathbf{b})$ is sufficiently large, the test statistic has an approximate chi-square distribution with d = *rank* $(\mathbf{D})$ degrees of freedom. However, if the degrees of freedom for $\hat{\mathrm{Var}}(\mathbf{b})$ is small (<60), then T_d^2 is distributed as a generalized T^2 statistic. That is,

$$F_{d,\, df+1-d} = \frac{df+1-d}{df*d} * T_d^2$$

has an F distribution with d and $df + 1 - d$ degrees of freedom, where df is the degrees of freedom for $\hat{\mathrm{Var}}(\mathbf{b})$.

4. Analyzing Logistic Regression Models With Survey Data Using WesVarPC

WesVarPC is a Windows-based software package developed at Westat, Inc. that computes estimates and replicate-based variance estimates from survey data collected using complex sampling and estimation procedures. WesVarPC reads a variety of input files and creates a WesVarPC file. Sampling errors are then estimated for different types of survey statistics.

WesVarPC supports a wide range of complex sample designs, including multistage, stratified, and unequal probability samples. The replicate variance estimation method can deal with many types of estimation schemes, such as poststratification or ratio estimation. Also, WesVarPC easily computes estimates of variances for complex functions of estimates including ratios, differences of ratios and log-odds ratios. Tests of hypotheses for tables can be performed using chi-square statistics that have been adjusted for the complex survey design. WesVarPC also computes parameter estimates for linear and logistic regression models. Along with providing a test for the overall fit of the regression model, WesVarPC can perform tests for the significance of linear combinations of variables included in the linear or logistic model.

The logistic model in WesVarPC computes estimates of the regession coefficients, the variance-covariance matrix of the estimated model parameters, and an R^2 statistic which is similar to the square of the multiple correlation coefficient (coefficient of determination) in the usual linear regession model. It also provides a test of the overall significance of the fitted logistic model, and the significance of linear combinations of parameters included in the model.

As indicated above, WesVarPC computes a Wald F statistic for the fit of a model. The F statistic is dependent on the number of degrees of freedom associated with the variance-covariance matrix of the parameters, $\hat{\mathrm{Var}}(\mathbf{b})$. The F statistic becomes less reliable when the degrees of freedom for $\hat{\mathrm{Var}}(\mathbf{b})$ is small (see Thomas and Rao 1987, and Korn and Graubard (1990)).

5. The Score Test for Logistic Regression Models

The score test provides an alternative procedure for testing hypotheses about model parameters. The score statistic is a quadratic form based on the vector of partial derivatives of the log-likelihood function with respect to the parameters of interest, evaluated at the values postulated by the null hypothesis. One important advantage of the score test relative to the Wald or Likelihood Ratio tests is computational efficiency, arising from the fact that the score does not require the computation of maximum likelihood estimates of the model parameters in testing for overall model significance.

The theoretical formulation of the score statistic and proof of its asymptotic equivalence to the Wald or Likelihood Ratio statistics under simple random sampling are now well established. See, for instance, Rao (1973), Cox and Hinkley (1974), Boos (1992), and references cited therein. In this section, we give a simple derivation of the score statistic for logistic regression models, and show how it can be applied to the analysis of complex survey data. We will then compare the properties of the score test with those of the Wald test via simulation using March 1995 U.S. Current Population Survey (CPS) data.

Following the notation given in earlier sections, the weighted likelihood function may be written as

$$\mathbf{L}(\beta|\mathbf{Y}) = \prod_{i\varepsilon S} P_i^{w_i Y_i}(1-P_i)^{w_i(1-Y_i)} = \prod_{i\varepsilon S}\left(\frac{P_i}{1-P_i}\right)^{w_i Y_i}(1-P_i)^{w_i}$$

where β, $\mathbf{Y}$, w_i, and P_i are as defined earlier and the product in the above equation is taken over all units i in the sample $S\,(i\varepsilon S)$. The log-likelihood function is

$$\begin{aligned} log_e\mathbf{L}(\beta\mid\mathbf{Y}) &= \sum_{i\varepsilon S}\left\{w_i log_e\left(\frac{P_i}{1-P_i}\right) + w_i log_e(1-P_i)\right\} \\ &= \sum_{i\varepsilon S} w_i Y_i \mathbf{X}_i'\beta - \sum_{i\varepsilon S} w_i log_e\left(1+e^{\mathbf{X}_i'\beta}\right) \end{aligned}$$

Therefore, the $(p+1)$ x 1 score vector, $S(\beta)$, is given by

$$S(\beta) = \frac{\partial}{\partial\beta'} log_e\mathbf{L}(\beta\mid\mathbf{Y}) = \sum_{i\varepsilon S} w_i\mathbf{X}_i'(Y_i - P_i)$$

which can be written in matrix notation as

$$S(\beta) = \mathbf{X}'\,\mathbf{W}(\mathbf{Y}-\mathbf{P})$$

Tests of hypotheses about model parameters will, in general, not include the intercept parameter β_0. Thus, testing the overall fit of a model corresponds to testing a hypothesis about the vector of parameters of interest: $\beta_1, \beta_2, \ldots, \beta_p$. If $V\{S(\beta)\}$ denotes the $(p+1) \times (p+1)$ covariance matrix of $S(\beta)$, then the score statistic for testing the null hypothesis $H_0: (\beta_1, \ldots, \beta_p) = (\beta_1^0, \ldots, \beta_p^0)$ versus H_1: not H_0 is given by

$$X_S^2 = [\mathbf{D}S(\beta^0)]'\left[\mathbf{D}V\left\{S(\beta^0)\right\}\mathbf{D}'\right]^{-1}[\mathbf{D}S(\beta^0)] \tag{4}$$

where $\mathbf{D}$ is a matrix of zeros and ones which is defined so that $\mathbf{D}\beta = (\beta_1, \ldots, \beta_p)$ and $\beta^0 = (\beta_0^0, \beta_1^0, \ldots, \beta_p^0)$. Under the null hypothesis, X_S^2 converges in distribution to a chi-square random variable with $d = rank\,(\mathbf{D})$ degrees of freedom.

The following examines a procedure for implementing the score test for testing hypotheses about parameters of interest when fitting logistic regression models to survey data using replication-based variance estimation methods. Computing the weighted score vector and the replicate-based estimate of its covariance matrix takes account of the complex sample design for this test. We shall consider tests of hypotheses about the whole parameter vector, as well as hypotheses about specified subsets of the parameter vector, using weights developed for the full sample and k replicate samples, as described in Section 3. The score statistic for testing the overall fit of the model is given by equation (4), where the covariance matrix of the full sample score vector is estimated by replication methods. For ease of exposition, we shall consider the full sample as one of the replicates, with index 0. Thus we have $k+1$ replicate weights, ranging from 0 to k, $(w_i^{(0)}, w_i^{(1)}, \ldots, w_i^{(k)})$, where replicate weight 0 is the full sample weight. First, for each replicate j, where, $j = 0, 1, \ldots, k$, we compute the $(p+1) \times 1$ score vector $S^j(\beta^0) = \mathbf{X}'\,\mathbf{W}^j(\mathbf{Y} - \mathbf{P}^0)$, where $S^0(\beta^0) = S(\beta^0)$ is the full sample score vector evaluated at β^0, $\mathbf{W}^j$ is the $n \times n$ diagonal matrix whose diagonal entries are provided by the j-th replicate weights, $\mathbf{P}^0 = (P_1^o, \ldots, P_n^o)$ is the vector of probabilities under the null hypothesis, and

$P_i^o = \left[1 + e^{-\mathbf{X}_i'\beta^0}\right]^{-1}$, $i = 1, 2, \ldots, n$.

The replication-based estimated covariance matrix of the score vector is then computed as

$$V\left\{S(\beta^0)\right\} = c\sum_{j=1}^{k}\left[S^j\left(\beta^{0(j)}\right) - S(\beta^0)\right]\left[S^j\left(\beta^{0(j)}\right) - S(\beta^0)\right]' \tag{5}$$

where c is as defined in Section 3. If the number of degrees of freedom associated with the estimated covariance matrix is large, then X_s^2 has an asymptotic chi-square distribution with $d = rank\,(\mathbf{D})$ degrees of freedom under H_o. Otherwise, X_s^2 can be converted to an F-statistic by an appropriate normalization, given by

$$F_s = \frac{df - d + 1}{df * d} X_S^2. \tag{6}$$

Under the null hypothesis, F_S has an asymptotic F distribution with d and $df - d + 1$ degrees of freedom, where df is the degrees of freedom associated with the estimated covariance matrix of the score vector. The maximum value for df is equal to k, the number of replicates. When the alternative hypothesis is true, F_S has an asymptotic non-central F distribution with d and $df - d + 1$ degrees of freedom, and noncentrality parameter (λ) given by

$$\lambda = [\mathbf{D}S(\beta^{(1)})]'\left[\mathbf{D}\hat{V}\left\{S(\beta^{(1)})\right\}\mathbf{D}'\right]^{-1}[\mathbf{D}S(\beta^{(1)})]$$

where $\beta^{(1)}$ is the vector of values of the model parameters under the alternative hypothesis.

We now extend the score test statistic to handle tests of hypotheses about a given subset of β, with a second subset treated as a vector of nuisance parameters. Let $\beta = (\mathbf{B}_1', \mathbf{B}_2')'$, where $\mathbf{B}_1$, and $\mathbf{B}_2$ are subvectors of β, of dimensions q and r respectively $(q + r = p + 1)$. Suppose we are interested in testing the hypothesis $H_o: \mathbf{B}_2 = \mathbf{B}_2^0$, while treating $\mathbf{B}_1$ as a vector of nuisance parameters. First, for each replicate j, where $j = 0, 1, \ldots, k$, we compute the maximum likelihood estimate $\hat{\mathbf{B}}_1^{0(j)}$ of $\mathbf{B}_1$ given that $\mathbf{B}_2 = \mathbf{B}_2^0$. Let $\hat{\beta}^{0(j)} = \begin{pmatrix}\hat{\mathbf{B}}_1^{0(j)} \\ \mathbf{B}_2^0\end{pmatrix}$. Next, we compute the score vector $S^j\left(\hat{\beta}^{0(j)}\right) = \mathbf{X}'\,\mathbf{W}^j\left(\mathbf{Y} - \hat{\mathbf{P}}^j\right)$, where $S^0\left(\hat{\beta}^{0(0)}\right) = S\left(\hat{\beta}^0\right)$ is the full sample score vector, $\hat{\mathbf{P}}^j = \left(\hat{P}_1^j, \ldots, \hat{P}_n^j\right)$ and $\hat{P}_i^j = \left[1 + e^{-\mathbf{X}_i'\hat{\beta}^{0(j)}}\right]^{-1}$, $i = 1, 2, \ldots, n$. We then compute the replicate-based covariance matrix $\hat{\mathrm{V}}\mathrm{ar}\left\{S\left(\hat{\beta}^0\right)\right\}$ of the full-sample score vector using equation (5). The score statistic for testing the above hypothesis is then given by (4), with $S(\hat{\beta}^0)$ in place of $S(\beta^0)$, $\hat{V}\{S(\hat{\beta}^0)\}$ in place of $V\{S(\beta^0)\}$, and where $\mathbf{D}$ is defined so that $\mathbf{D}\beta = \mathbf{B}_2$. Under H_o, $X_{S(2)}^2$ has an asymptotic chi-square distribution with $d = rank(\mathbf{D})$ degrees of freedom.

5.1 Simulation Study

Using March 1995 CPS data, a simulation study was designed to conduct a preliminary evaluation of the performance of the score test for testing hypotheses about parameters in logistic regression models as compared to the Wald-F test, which is currently being used by WesVarPC. The CPS File used for the simulation study consisted of about 114,000 observations. These observations were considered the entire population for the simulation (not a sample of the US population).

5.1.1 The Logistic Regression Model

For the simulation study, we used a logistic regression model for predicting the propensity (P_i) for an individual to be on public assistance. The three auxiliary variables used in the model are X_1 (actual age in years); X_2 (in two categories: 1 if children are present in the household, and 0 otherwise); and X_3 (in two categories: 1 if home is owned, and 0 otherwise). The model is

$$log_e\left(\frac{P_i}{1-P_i}\right) = \beta_0 X_{0i} + \beta_1 X_{1_i} + \beta_2 X_{2_i} + \beta_3 X_{3_i} + \beta_4 X_{4_i} + \varepsilon_i$$

where $X_0 = 1$; $X_4 = X_1 * X_3$; β_0, β_1, β_2, β_3, β_4 are the model parameters; and ε_i is the random error associated with the i-th sampled individual. Three stratification designs and two sample sizes were used in this simulation study. A two primary sampling unit (PSU) per stratum design, which included 16, 32, and 64 strata, was used. Two hundred samples of size of about 1,000 and 5,000 each were selected from the above stratified designs, with equal numbers of records selected from each PSU. For instance, eight records were selected from each PSU in the 64 strata design for a total of 1,024 sampled cases. The following hypotheses were tested for each sample, using the procedures described above:

$$H_{0(1)}: (\beta_1, \beta_2, \beta_3, \beta_4) = (0, 0, 0, 0)$$
$$H_{0(2)}: \beta_4 = 0$$
$$H_{0(3)}: (\beta_1, \beta_4) = (0, 0)$$

For each test, the score-based F statistic (Score-F), the Wald-F statistic, and their respective p-values were calculated for all 200 samples.

5.1.2 Empirical Results

For each sample and each set of replicate weights, the three hypotheses above were tested at three different significant levels ($\alpha = 0.01$, 0.05, and 0.10). Table 1 gives a comparison of the rejection rates (the proportion of times $H_{0(i)}$, i = 1, 2, 3, was rejected over repeated sampling) for the Score-F and Wald-F tests for testing each of the hypotheses under consideration.

The hypothetical true values of the model parameters ($(\beta_0$, β_1, β_2, β_3, $\beta_4)$ = (-0.9414, -0.0282, -1.5877, -1.4808, 0.0098), were obtained by fitting the model to the entire population using PROC LOGISTIC in SAS.

Based on the values of the model parameters, $H_{0(1)}$ and $H_{0(3)}$ are false but $H_{0(2)}$ is almost true. Thus, for $H_{0(2)}$, Table 1 entries provide values close to estimated significance levels for the various sample sizes. For $H_{0(1)}$ and $H_{0(3)}$, Table 1 entries are equal to the proportion of times a false hypothesis is rejected. For all three tests, both Score-F and Wald-F perform well with sample size of 5,120 and 62 replicates (the maximum degrees of freedom). In general, the two tests are somewhat similar in performance.

The results obtained in the preliminary empirical study illustrate the potential application of score test for logistic models. The present findings, while quite informative, are certainly not conclusive. Further research is needed to evaluate the performance of the score test under various conditions and especially when the degrees of freedom associated with the statistic is small. Future research will include comparisons of the distribution functions of score test and Wald-F test statistics when the null hypothesis is not true.

Acknowledgments

This research was initiated after discussions with J.N.K. Rao on the application of score test to data from complex surveys. We are grateful for the discussions and the initial motivation. Rao's relevant research on score test was presented at the 1996 Canadian Statistical Society annual meetings and will appear in the proceedings ("Fitting Models to Survey Data" by J.N.K. Rao and A.J. Scott).

REFERENCES

Boos, D.B. (1992). On Generalized Score Test. *The American Statistician*, 46, 4, pp. 327-333.

Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics*. London: Chapman and Hall.

Kleinbaum, D.C., Kupper, L.L., and Morgenstern, H. (1982). *Epidemiologic Research, Principles and Quantitative Methods*, Lifetime Learning Publications, A Division of Wadsworth, Inc., Belmont, California.
Korn, E.L., and Graubard, B.I. (1990). "Simultaneous Testing of Regression Coefficients with Complex Survey Data: Use of Beneferroni t-statistic," The American Statistician.
Hosmer, D.W., and Lemeshow, S. (1989). "Applied Logistic Regression," Wiley.
McCarthy, Philip J. (1969). "Pseudo-Replication: Half-Samples," *Review of International Statistical Institute,* 37, 239-264.
Pfeffermann, D. (1993). "The Role of Sampling Weights When Modeling Survey Data," International Statistical Review 61, 2, pp. 317-337.
Rao, C.R. (1973). *Linear Statistical Inference and its Applications*, Second Edition, New York: John Wiley.
Sarndal, C.E., Swenson, B., and Wretman, J. (1992). "Model Assisted Survey Sampling," Springer-Verlag.
Thomas, D.R., and Rao, J.N.K. (1987). "Small Sample Comparisons of Levels and Power for Simple Goodness-of-Fit Statistics Under Cluster Sampling," Journal of the American Statistical Association, 82, pp. 630-636.
Wolter, K.M. (1985). Introduction to Variance Estimation. New York: Springer-Verlag.

Table 1. Proportion of times null hypothesis is rejected

Hypothesis	Sample size	Number of replicates	Test statistic	Significance level 0.01	0.05	0.10
$H_{0(1)}$	1,024	16	Score-F	0.83	0.97	0.98
			Wald-F	0.83	0.93	0.97
		32	Score-F	0.90	0.97	0.98
			Wald-F	0.85	0.96	0.98
		64	Score-F	0.89	0.97	0.99
			Wald-F	0.89	0.95	0.96
	5,120	16	Score-F	1.00	1.00	1.00
			Wald-F	1.00	1.00	1.00
		32	Score-F	1.00	1.00	1.00
			Wald-F	1.00	1.00	1.00
		64	Score-F	1.00	1.00	1.00
			Wald-F	1.00	1.00	1.00
$H_{0(2)}$	1,024	16	Score-F	0.01	0.01	0.02
			Wald-F	0.03	0.11	0.18
		32	Score-F	0.00	0.01	0.02
			Wald-F	0.05	0.09	0.16
		64	Score-F	0.00	0.00	0.01
			Wald-F	0.05	0.13	0.17
	5,120	16	Score-F	0.01	0.03	0.09
			Wald-F	0.02	0.04	0.06
		32	Score-F	0.03	0.06	0.13
			Wald-F	0.03	0.08	0.12
		64	Score-F	0.01	0.05	0.09
			Wald-F	0.01	0.04	0.08
$H_{0(3)}$	1,024	16	Score-F	0.38	0.63	0.75
			Wald-F	0.25	0.55	0.67
		32	Score-F	0.40	0.66	0.77
			Wald-F	0.32	0.48	0.90
		64	Score-F	0.37	0.63	0.79
			Wald-F	0.29	0.54	0.65
	5,120	16	Score-F	0.98	1.00	1.00
			Wald-F	0.94	0.99	1.00
		32	Score-F	0.98	1.00	1.00
			Wald-F	0.96	0.99	1.00
		64	Score-F	0.99	1.00	1.00
			Wald-F	0.96	0.99	1.00