Computational Statistics & Data Analysis 5 (1987) 201–213
North-Holland

The effect of imputed values on the distribution of the goodness-of-fit chi-square statistic

Phyllis A. GIMOTTY

Department of Statistics, Division of Biostatistics, University of Florida, Gainesville, FL 32610, USA

Morton B. BROWN

Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA

Received 13 December 1985
Revised 27 February 1987

Abstract: A method used to compensate for nonresponse is to impute missing values; that is, to replace each missing value with a respondent value selected from all observed values or from a subset of observed values. The imputation procedure used in this paper selects imputed values from the respondent data using simple random sampling with replacement within homogeneous subsets and replaces the missing values with these values to complete the data set. The empirical distribution of the goodness-of-fit chi-square statistic computed from the 'completed' data set is compared to its asymptotic distribution and to the distribution of the traditional chi-square test statistic applied to the completed data set by ignoring the imputation.

At nominal levels of five and ten percent, the asymptotic distribution of the goodness-of-fit chi-square statistic computed from the completed data set is shown to have a good empirical behavior at moderate sample sizes. When the imputed values are treated as actual responses and the imputation is ignored, the empirical levels of significance are much larger than the nominal levels.

Keywords: Missing data, Imputation, Categorical data, Resampling plans.

1. Introduction

Many data sets contain observations for some individuals in which the data are incomplete. When the characteristics of the individuals who do not respond fully are different in distribution from those of the entire population, statistics computed from the respondent data may be biased estimates of the population parameters.

Imputation procedures, such as those used by the U.S. Bureau of the Census, replace values that are missing with observed responses, within homogeneous subsets of the data that are referred to as imputation classes. (For examples see Bailar, Bailey and Corby [1] and Nisselson [8].)

This technique produces a 'completed' or 'filled-in' data set which has a value recorded for each individual in the data set. Completed data sets are available to researchers in the form of public use data tapes for some national surveys. If the data are analyzed along traditional lines, the imputed, or predicted values, are treated as real. This paper will consider statistical inferential techniques that take into account the special character of these data sets.

Proportions computed from such data sets are potentially less biased estimates of population characteristics than estimates computed from the respondent data alone (Ford [6]). However, the covariance structure of these proportions is affected since the imputed values are correlated with the observed responses. As a consequence, statistics computed from these proportions have a more complicated distribution theory than proportions computed from a data set with complete observations. For example, the goodness-of-fit chi-squared statistic no longer has the traditional chi-squared distribution for a class of imputation procedures (Gimotty [4]).

In this paper we investigate the imputation procedure that replaces each missing value with a respondent value selected from all the responses in the appropriate imputation class by simple random sampling with replacement. Simulated data are used to investigate how well the asymptotic distribution of the goodness-of-fit chi-square statistic calculated from proportions computed from the completed data set approximates its empirical distribution at moderate sample sizes. A test for goodness of fit using an estimated asymptotic distribution function is proposed and the empirical size of the test for goodness of fit derived from the asymptotic distribution of the chi-square statistic computed from a completed data set is compared with the empirical size of the usual chi-square test for goodness of fit when the imputed values are treated as if they were actual responses.

2. Test for goodness of fit

The categorical response variables for the units in the population are assumed to have independent, but not necessarily identical, multinomial distributions. A simple random sample of n units is selected from the finite population. Within each of C imputation classes, the responses are assumed to have a multinomial distribution with parameter vector $\boldsymbol{\theta}_k$ $(k = 1, 2, \ldots, C)$ which can differ between imputation classes. The vector of population probabilities for the response variable for any random unit in the population is a weighted sum of the imputation class parameters

$$\boldsymbol{\pi} = \sum_{k=1}^{C} \eta_k \boldsymbol{\theta}_k$$

where η_k is the probability of being in the k-th imputation class. The vector of

population probabilities for the response variable for any random responding unit in the population is defined by

$$\pi_R = \sum_{k=1}^{C} \left(\frac{\rho_k}{\rho} \right) \theta_k$$

where ρ_k is the probability of a unit responding in the k-th imputation class. The probability of response, ρ, is defined by

$$\rho = \sum_{k=1}^{C} \rho_k .$$

A measure of goodness-of-fit of the sample proportions to the population probabilities is Pearson's chi-square statistic, X_n^2, that is,

$$X_n^2 = n \sum_{i=1}^{I} \frac{(p_{in} - \pi_i)^2}{\pi_i}$$

where $\boldsymbol{p}_n = (p_{1n}, p_{2n}, \ldots, p_{In})$ is the vector of proportions in each of the I categories of the response variable. When proportions computed from a completed data set are used to calculate this statistic, it no longer has a simple asymptotic chi-squared distribution with $I - 1$ degrees of freedom.

Gimotty [4] derived the asymptotic distribution of the chi-square statistic computed from proportions obtained from a completed data set for a general class of imputation procedures. Under the null hypothesis that π are the underlying population probabilities, its asymptotic cumulative distribution function is $1 - H(x \mid \lambda)$ where $H(x \mid \lambda) = \Pr(X_n^2 > x)$ is given by

$$H(x \mid \lambda) = \frac{1}{2} + \frac{1}{\pi} \int_0^{\infty} \frac{\sin[\theta(u, x)]}{u \cdot \delta(u)} \, du, \tag{1}$$

such that

$$\theta(u, x) = \tfrac{1}{2} \sum_{i=1}^{I} \tan^{-1}(\lambda_i u) - \tfrac{1}{2} u x,$$

$$\delta(u) = \prod_{i=1}^{I} \left[1 + (\lambda_i u)^2 \right]^{1/4},$$

where λ are the eigenvalues of the matrix $D(\pi)^{-1}\Sigma$, with $D(\pi)$ a diagonal matrix whose diagonal elements are given by the vector π, and with Σ the conditional asymptotic covariance matrix of the vector of proportions computed from the completed data set. When the imputation procedure uses simple random sampling with replacement within imputation classes, the covariance of the vector of proportions computed from the completed data set, Σ, is given by

$$\Sigma = \sum_{k=1}^{C} \omega_k \left[D(\boldsymbol{\theta}_k) - \boldsymbol{\theta}_k \boldsymbol{\theta}_k' \right] \tag{2}$$

such that

$$\omega_k = \frac{\rho_k}{\rho^2}\left[1-(1-\rho)^2\right] + \frac{(n_k-\rho_k)^2}{\rho_k}\left[1+\frac{\rho_k}{\eta_k-\rho_k}\right], \tag{3}$$

where ρ_k is the probability of a unit responding in the k-th imputation class, η_k is the probability of observing the k-th imputation class, and ρ is the probability of response. This distribution is a constant times a chi-squared distribution with $I-1$ degrees of freedom when the imputation class parameters, θ_k, are all equal to a common value π.

The expected value of the chi-square statistic is given by

$$E\left[X_n^2\right] = \sum_{k=1}^{C} \omega_k \sum_{i=1}^{I} \frac{\theta_{ik}(1-\theta_{ik})}{\pi_i} \tag{4}$$

where ω_k is defined by (3). When all of the imputation class parameters are equal to a common value, π, this expectation simplifies to

$$E\left[X_n^2\right] = (I-1)\sum_{k=1}^{C} \omega_k.$$

In general, the expected value is sensitive to differences in the probability of response (ρ_k), differences between the probability of response in each imputation class and the probability of each imputation class ($\rho_k - \eta_k$) and differences between the imputation class parameters and the population parameters $[\theta_{ik}(1-\theta_{ik})/\pi_i]$ as well as the relationships between these quantities. The degrees of freedom of the chi-square statistic computed from complete data, $I-1$, is a lower bound for the expected value only when the imputation class parameters and response rates are all equal to a common value.

The variance of the chi-square statistic is given by

$$V\left[X_n^2\right] = 2\left[\sum_{i=1}^{I} \frac{\left[\sum_{k=1}^{C} \omega_k \pi_{ik}(1-\pi_{ik})\right]^2}{\pi_i^2} + \sum_{i\neq j} \frac{\left[\sum_{k=1}^{C} \omega_k \pi_{ik}\pi_{jk}\right]^2}{\pi_i \pi_j}\right] \tag{5}$$

and when $\pi_k = \pi$ for each imputation class, this simplifies to

$$V\left[X_n^2\right] = 2(I-1)\left(\sum_{k=1}^{C} \omega_k\right)^2.$$

The categorical response variables are independent and identically distributed random variables when $\pi_k = \pi$ for all imputation classes. Consequently, the respondent proportions as well as the proportions computed from the completed data set are unbiased estimators. The asymptotic distribution of X_n^2 calculated using the proportions computed from the completed data set is a constant times a chi-squared distribution with $I-1$ degrees of freedom. However, the imputation

procedures are used because the proportions computed from the respondent data are thought to be biased since the distribution of the categorical response variable and the response rate is thought to be different for each imputation class. In these situations, simple procedures, such as rescaling the chi-square statistic and using the chi-squared distribution with $I-1$ degrees of freedom, are not good approximations to the asymptotic distribution of the chi-square statistic computed from a completed data set [5].

When the underlying parameters are known, the test for goodness of fit would reject the hypothesis that π is the underlying vector of population probabilities when the attained significance value of the chi-square statistics, $H(X_n^2 \mid \lambda)$, is less than or equal to α where α is the nominal level of the test. However, in practice, there is generally insufficient information to specify hypothetical values for these parameters for each imputation class.

We propose to estimate the covariance matrix of the proportions, Σ, by using consistent estimates of the nuisance parameters. The distribution function is approximated by $H(x \mid \hat{\lambda})$ where $\hat{\lambda}$ are estimates of the eigenvalues of the matrix $D(\pi)^{-1}\hat{\Sigma}$ and $\hat{\Sigma}$ is the consistent estimate of the covariance matrix of the vector of proportions computed from the completed data set. The matrix $\hat{\Sigma}$ is obtained from (2) by substituting consistent estimates of the nuisance parameters given by $\hat{\rho} = r/n$, $\hat{\rho}_k = r_k/n$, $\hat{\eta}_k = n_k/n$ and $\hat{\boldsymbol{\theta}}_k = \boldsymbol{p}_{kr}$ where r is the number of respondents in n sampled units, r_k and n_k are the number of respondents and sampled units in the k-th imputation class, respectively, and $\boldsymbol{p}_{kr}$ is the vector of proportions computed from the respondent data in the k-th imputation class. The consistency of these estimates follows from the properties of simple random sampling. The proposed test for goodness of fit rejects the hypothesis that π is the vector of population probabilities when $H(X_n^2 \mid \hat{\lambda}) < \alpha$ where α is the nominal level of the test.

3. The simulation

The goodness-of-fit test is useful only if the distribution of the imputed chi-square statistic is adequately approximated by its estimated asymptotic distribution at sample sizes used in practice. The objective of the simulation was to identify conditions where this approximation failed at moderate sample sizes. In this section we describe the simulation used to study the empirical behavior of the goodness-of-fit chi-square statistic. These data are used to investigate the moments of the chi-square statistics and the empirical sizes for the goodness-of-fit test using the chi-square statistic computed from a completed data set. The results of the simulation are presented in Section 4.

3.1. Parameterization of the model

Each data set generated in the simulation is a realization of a three-dimensional contingency table defined by: the response variable, X, with I categories;

the imputation class variable, Z, with C categories; and the indicator variable, R, that a unit responds. Data are assumed to be missing at random (Rubin [9]) within imputation classes. As a consequence, the response variable and the indicator variable for a responding unit are conditionally independent given the imputation class. The probability that a unit is in category i in the k-th imputation class for event j (nonresponse or response) is Δ_{ijk} and it can be expressed as

$$\Delta_{ijk} = \Pr[X = i \wedge R = j \mid Z = k] \cdot \Pr[Z = k]$$
$$= \Pr[X = i \mid Z = k] \cdot \Pr[Z = k \mid R = j] \cdot \Pr[R = j].$$

Therefore, the model defining these probabilities is specified by the vectors: $\boldsymbol{\theta}_k$, the imputation class parameters ($k = 1, 2, \ldots, C$); $\boldsymbol{\kappa}_1$ and $\boldsymbol{\kappa}_0$, the vectors of imputation class probabilities for the respondents or nonrespondents, respectively; and ρ, the overall probability of response.

A model for the probabilities $\boldsymbol{\pi}_k$ was selected to provide a range of possible distributions as well as to provide a simple way of describing the relationship between the different imputation classes. Six categories were used for the categorical response variable, x, as well as six imputation classes, Z, to provide symmetry. A distribution was selected and the probabilities were ordered from largest value to smallest to define $\boldsymbol{\theta}_1$; the remaining parameter vectors are defined by permutations of $\boldsymbol{\theta}_1$. For independence, $\boldsymbol{\theta}_k = I\boldsymbol{\theta}_1$ ($k = 2, 3, \ldots, 6$), where I is the 6×6 identity matrix and for dependence with positive association, $\boldsymbol{\theta}_k = A\boldsymbol{\theta}_{k-1}$ ($k = 2, 3, \ldots, 6$), where A is given by $(\delta_2', \delta_3', \ldots, \delta_I', \delta_1')$ and δ_i is a $1 \times C$ vector whose i-th element is 1 and all other elements are zero. Models with negative association were not considered.

3.2. Models used in the simulation

The first models investigated were defined using the uniform distribution and four response rates, 0.90, 0.75, 0.60, and 0.45. The remaining models were selected by varying only one dimension of the parameter space with a response rate of 0.60. These models include situations where the probabilities for the response categories differed between imputation classes and relationships other than independence exist between the imputation class variable and both the response variable and the indicator variable for a responding unit. More importantly, these are models where the distribution for the respondents differed from the distribution of the nonrespondents. All but the first model in Table 1 have an extreme value for one of the parameters of the model and in these situations the empirical distribution is more likely to deviate from the asymptotic theory.

Table 1 summarizes the models used in the simulation. The imputation class parameters, $\boldsymbol{\theta}_1$, and the imputation class probabilities for the respondents, $\boldsymbol{\kappa}_1$, define each model when the response rate is 0.60. The categorical response variable and the imputation class variable are independent for models 1–4 and 6

Table 1
Models used in the simulation of the distribution of X_n^2 when $\rho = 0.60$

Parameter	Model														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
θ_{11}	0.167	0.300	0.300	0.500	0.300	0.300	0.300	0.500	0.140	0.080	0.120	0.300	0.300	0.080	0.048
θ_{12}	0.167	0.140	0.140	0.100	0.140	0.140	0.140	0.100	0.151	0.115	0.139	0.140	0.140	0.115	0.095
θ_{13}	0.167	0.140	0.140	0.100	0.140	0.140	0.140	0.100	0.161	0.149	0.157	0.140	0.140	0.149	0.143
θ_{14}	0.167	0.140	0.140	0.100	0.140	0.140	0.140	0.100	0.172	0.184	0.176	0.140	0.140	0.184	0.190
θ_{15}	0.167	0.140	0.140	0.100	0.140	0.140	0.140	0.100	0.183	0.219	0.195	0.140	0.140	0.219	0.238
κ_{11}	0.167	0.167	0.047	0.167	0.047	0.047	0.167	0.167	0.213	0.213	0.286	0.047	0.286	0.286	0.286
κ_{12}	0.167	0.167	0.095	0.167	0.095	0.095	0.167	0.167	0.195	0.195	0.230	0.095	0.230	0.230	0.230
κ_{13}	0.167	0.167	0.143	0.167	0.143	0.143	0.167	0.167	0.176	0.176	0.190	0.143	0.190	0.190	0.190
κ_{14}	0.167	0.167	0.190	0.167	0.190	0.190	0.167	0.167	0.157	0.157	0.143	0.190	0.143	0.143	0.143
κ_{15}	0.167	0.167	0.238	0.167	0.238	0.238	0.167	0.167	0.139	0.139	0.095	0.238	0.095	0.095	0.095
π_1	0.167	0.300	0.300	0.500	0.148	0.300	0.167	0.167	0.167	0.167	0.167	0.163	0.170	0.167	0.168
π_2	0.167	0.140	0.140	0.100	0.155	0.140	0.167	0.167	0.166	0.166	0.166	0.164	0.169	0.165	0.164
π_3	0.167	0.140	0.140	0.100	0.163	0.140	0.167	0.167	0.166	0.165	0.165	0.166	0.167	0.164	0.162
π_4	0.167	0.140	0.140	0.100	0.171	0.140	0.167	0.167	0.166	0.166	0.166	0.167	0.166	0.165	0.164
π_5	0.167	0.140	0.140	0.100	0.178	0.140	0.167	0.167	0.167	0.167	0.167	0.169	0.164	0.167	0.168
$\|\pi - \pi_R\|(\times 10^2)$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.013	0.024	0.065	0.065	0.086	0.162
$\min_k \|\theta_k - \pi\|$	0.000	0.000	0.000	0.000	0.016	0.016	0.021	0.133	0.002	0.020	0.006	0.023	0.023	0.020	0.037
$\max_k \|\theta_k - \pi\|$	0.000	0.000	0.000	0.000	0.028	0.028	0.021	0.133	0.002	0.020	0.007	0.020	0.020	0.023	0.043
$E[X_n^2]$	10.3	10.3	12.6	10.3	10.1	10.1	10.1	8.6	10.5	10.3	12.5	12.3	12.3	12.3	12.0
$V[X_n^2]$	42.7	42.7	63.7	42.7	40.7	40.7	40.6	30.1	44.2	42.3	62.9	60.7	60.7	60.8	58.3

and they have a positive association otherwise. The Euclidean distance between the vector of population probabilities and the vector of probabilities for the responding units, $\| \pi - \pi_R \|$, is a measure of the difference between the distribution for the response variable, X, for the respondents and the nonrespondents. The minimum and maximum Euclidean distances of the imputation class parameter vectors, θ_k, and the vector of population probabilities, π, are a measure of heterogeneity of the imputation class parameters relative to the population probabilities. The expected value and variance found in Table 1 are the asymptotic expected value and variance for the chi-square statistic computed from the completed data set when the imputation procedure uses simple random sampling with replacement within imputation classes given by (4) and (5), respectively.

The first model describes a situation where the distribution for both the categorical response variable and the imputation class variable is the uniform distribution and all three categorical variables are independent; the next seven models are situations where the population probabilities are equal to the population respondent probabilities, $\| \pi - \pi_R \| = 0$. For these eight models is the nonrespondents are expected to respond like the respondents. However, for the last seven models $\| \pi - \pi_R \| > 0$ and the nonrespondents are expected to respond differently from the respondents.

Of the eight models where the respondents were expected to respond like the nonrespondents, there were six models where the probability distribution for the imputation class variable was uniform. In this situation, when the sample size is 500, the expected number of units in the sample in each of the imputation classes was 83. For the two models where the distribution of the imputation class variable was not uniform, the differences between adjacent probabilities were equal and the expected numbers of observations in each imputation class were 24, 48, 71, 95, 119 and 143. For two of the remaining seven models, where the respondents and nonrespondents have a different distribution, the expected numbers of observations in each imputation class were 88, 86, 84, 82, 81, and 79; for the other five models, the expected numbers of observations in each imputation class were 95, 90, 86, 81, 76, and 71.

3.3. The algorithms

The asymptotic distribution depends on the variance of the proportions computed from a completed data set conditional on the number of units and responding units within the sample in each imputation class. However, rather than generating data for each contingency table with these marginal distributions fixed, data for each contingency table were generated randomly. This simplified the design of the simulation.

One thousand contingency tables were generated where each table had 500 observations. These tables are combined in pairs to create 500 contingency tables with 1000 observations and in quadruplets to create 250 tables with 2000 observations. The information on the value for the response variable, X, was set to missing for the nonrespondents; nonrespondents in each imputation class were

assigned imputed values using simple random sampling with replacement from the respondents in the nonrespondent's imputation class to create the completed data set. The goodness-of-fit chi-square statistic was computed using each of these completed data sets.

The generation of the contingency tables, imputation of missing values and calculations of the chi-square statistics were done on an IBM XT Personal Computer. The descriptive statistics and the calculation of the attained significance values from the chi-square statistics were done in The University of Michigan's Amdahl computer.

The algorithm used to generate the uniform random numbers was based on the random number generator used by the BMDP statistical software package [2]. The alias method [7] was used to generate the contingency tables. The imputed values were randomly generated from a multinomial distribution whose parameter was the vector of respondent proportions using the table look-up algorithm for each imputation class. An algorithm by Sheil and O'Muircheartaigh [10] was used to numerically approximate the distribution function for which eigenvalues were calculated using a subroutine from the EISPACK library [3].

4. Results

First we consider the model where the probabilities for the response variable and for the imputation class variable have a uniform distribution. The means, standard deviations and asymptotic expected values of the goodness-of-fit chi-square statistics computed from the T completed data sets and the T respondent

Table 2
Mean and standard deviation of the chi-square statistics for model 1 at different response rates and sample sizes

Response rate	Sample size						Expected value	Standard deviation
	$n = 500$		$n = 1000$		$n = 2000$			
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.		
	Respondent Data							
0.90	4.99	3.27	5.01	3.32	5.16	3.10	5.00	3.16
0.75	4.98	3.16	5.09	3.49	5.25	3.55	5.00	3.16
0.60	5.08	3.38	5.16	3.52	5.14	3.09	5.00	3.16
0.45	4.99	3.21	4.99	3.24	5.15	3.30	5.00	3.16
	Completed Data							
0.90	6.05	3.90	6.05	4.14	6.23	3.59	6.06	3.38
0.75	7.80	4.91	7.96	5.52	8.31	5.48	7.92	5.01
0.60	10.65	6.87	10.76	6.91	10.62	6.94	10.33	6.54
0.45	14.01	8.96	14.10	8.77	14.46	9.44	13.86	8.77
	$T = 1000$		$T = 500$		$T = 250$			

Table 3
Estimated tail probabilities of the chi-square statistic for model 1 at different response rates and sample sizes

Response rate	Tail probabilities computed using known distribution function			Tail probabilities computed using estimated distribution function		
	$n = 500$	$n = 1000$	$n = 2000$	$n = 500$	$n = 1000$	$n = 2000$
	Nominal Level = 0.05					
0.90	0.047	0.036	0.056	0.052	0.038	0.060
0.75	0.045	0.066	0.072	0.047	0.066	0.076
0.60	0.057	0.064	0.068	0.060	0.064	0.072
0.45	0.049	0.052	0.068	0.050	0.050	0.072
	Nominal Level = 0.10					
0.90	0.098	0.096	0.108	0.099	0.100	0.104
0.75	0.095	0.108	0.136	0.088	0.106	0.132
0.60	0.113	0.114	0.120	0.127	0.124	0.124
0.45	0.098	0.106	0.140	0.112	0.118	0.140
	$T = 1000$	$T = 500$	$T = 250$	$T = 1000$	$T = 5000$	$T = 250$

data sets are shown in Table 2 for four different response rates and three sample sizes (n). For both goodness-of-fit chi-square statistics computed from the completed data and the respondent data alone, the ninety-five percent confidence intervals computed from the mean and standard deviation contained the asymptotic expected value.

Table 4
Mean and standard deviation of the chi-square statistics computed from the respondent data set and the completed data set with a response rate of 0.60 and a sample size of 500 ($T = 1000$)

Model	Respondent data			Completed data		
	Mean	Std. dev.	Exp. val.	Mean	Std. dev.	Exp. val.
1	5.08	3.38	5.00	10.65	6.87	10.33
2	5.15	3.22	5.00	10.45	6.45	10.33
3	4.95	3.10	5.00	12.62	8.32	10.62
4	5.02	3.36	5.00	10.33	6.55	10.33
5	5.06	3.03	5.00	10.21	6.15	10.09
6	4.89	3.01	4.88	10.57	6.07	10.33
7	5.07	3.18	4.87	10.38	6.35	10.07
8	5.07 [a]	3.03	4.20	9.54 [a]	5.85	8.68
9	4.86	2.96	5.01	10.32	6.53	10.52
10	5.28	3.36	5.11	10.60	6.71	10.28
11	5.59	3.65	5.41	13.33 [a]	8.20	12.53
12	6.19	3.71	6.04	12.89 [a]	7.85	12.30
13	6.01	3.64	6.04	12.64	8.05	12.30
14	6.67 [a]	3.94	6.39	13.00 [a]	8.22	12.31
15	7.71	4.46	7.62	12.24	8.00	12.04

[a] Significant with $p \leqslant 0.05$.

The empirical sizes of the goodness-of-fit test using the asymptotic distribution with both known and estimated parameters at the nominal five and ten percent levels of significance are presented in Table 3. All of the estimated tail probabilities are consistent with the nominal level of the test.

The means and standard deviations obtained from 1000 chi-square statistics computed from the completed data sets with 500 recorded values are presented in Table 4 for all models described in Table 1 where the response rate is 0.60. The expected values presented in this table are calculated using (4). In most of these situations, the ninety-five percent confidence intervals contained the asymptotic expected value.

The asymptotic distribution of the chi-square statistic was derived using the variance of the proportions computed from the completed data set conditional on the number of units in each imputation class. The asymptotic expected value of the chi-square statistic including the component of variance due to the randomness associated with the proportions in each imputation class is

$$E_u[X_n^2] = E[X_n^2] + \sum_{k=1}^{C} \eta_k \sum_{i=1}^{I} \frac{(\theta_{ik} - \pi_i)^2}{\pi_i}$$

where $E[X_n^2]$ is given by (4). When the sum $\Sigma\eta_k\Sigma[(\theta_{ik} - \pi_i)^2/\pi_i]$ is large, the empirical distribution function may deviate from the asymptotic distribution defined by (1). There are two such examples in Table 4 (Models 8 and 14); these are models where the chi-square statistics computed using the completed data set and the respondent data are significantly different from the asymptotic expected value.

Table 5
Estimated tail probabilities ($T = 1000$)

Model	Distribution function				
	$H(X\|\lambda)$		$H(X\|\hat{\lambda})$		$\chi^2(5)$
	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$
1	0.057	0.113	0.060	0.127 [a]	0.383
2	0.052	0.100	0.049	0.107	0.369
3	0.046	0.088	0.055	0.098	0.496
4	0.061	0.110	0.056	0.115	0.370
5	0.048	0.091	0.052	0.102	0.340
6	0.058	0.103	0.055	0.101	0.377
7	0.053	0.102	0.057	0.104	0.373
8	0.066 [a]	0.130 [a]	0.075 [a]	0.134 [a]	0.341
9	0.043	0.077 [a]	0.046	0.083	0.376
10	0.059	0.109	0.061	0.138 [a]	0.370
11	0.063	0.114	0.064	0.120	0.533
12	0.053	0.113	0.049	0.113	0.520
13	0.060	0.103	0.062	0.115	0.506
14	0.066 [a]	0.127 [a]	0.073 [a]	0.150 [a]	0.510
15	0.058	0.112	0.063	0.116	0.481

[a] Significant with $p \leq 0.05$.

The empirical sizes of the goodness-of-fit test using both the estimated and the known distribution function are presented in Table 5. In general, the estimated tail probabilities are consistent with the nominal level of the test; however, they have a slight positive bias. The mean difference between the attained significance values computed using the estimated and the known distribution function is 0.0044 with a standard deviation of 0.0023 when the nominal level is 0.05 and is 0.0088 with a standard deviation of 0.0079 when the nominal level is 0.10.

The estimated tail probabilities of the chi-square statistic when the imputed values are treated as actual responses are also given in Table 5 when the nominal level is 0.05. In this case the chi-square statistics computed from a completed data set are treated as random variables that have a chi-squared distribution with $I - 1$ degrees of freedom. The results indicate that these estimated tail probabilities severly overestimate the five percent nominal level of the test, in some caes by a factor of ten.

5. Discussion

This simulation demonstrates that treating imputed values as actual response values and using methodology developed for a complete data set results in severe bias of the estimated tail probabilities. Modifications to traditional methodology to compensate for the imputed values are necessary when completed data sets are used for inference.

The empirical means of the goodness-of-fit chi-square statistic computed from a completed data set are consistent with the asymptotic expected value in most cases. The empirical size of the goodness-of-fit test defined using the asymptotic distribution of the chi-square statistic computed from a completed data set at both nominal levels of 0.05 and 0.10 recommends the use of this test even at moderate sample sizes.

For a wide range of situations, the asymptotic approximations are valid when the number of observations in each imputation class is at least seventy and may be valid when one or two of the imputation classes have as few as twenty-five observations. In addition, the bias of the attained significance levels calculated using the estimated distribution function is not practically significant in the tails of the distribution. The use of this test for goodness-of-fit should be limited to situations where $\Sigma\eta_k\Sigma[(\theta_{ik} - \pi_i)^2/\pi_i]$ is small; a restriction which is generally satisfied in practice.

Acknowledgements

This research was done in partial fulfillment on the first author's Ph.D. dissertation done at the University of Michigan, Department of Biostatistics. The preparation of this paper was supported by U.S. National Institute for Mental Health grant number MH-37188.

References

[1] B.A. Bailar, L. Bailey, and C. Corby, A comparison of some adjustment and weighting procedures for survey data, in: N.K. Namboodiri (Ed.), *Survey Sampling and Measurement* (Academic Press, New York, 1978) 175–198.

[2] W.J. Dixon (Ed.), *BMDP Statistical Software* (University of California Press, Berkeley, CA, 1983).

[3] B.S. Garbow, J.M. Boyle, J.J. Dongarra, and C.B. Moler, Matrix eigensystem routines – EISPACK guide extension, *Lecture Notes in Computer Science*, Vol. 51 (Springer, Berlin, 1977).

[4] P.A. Gimotty, Goodness-of-fit chi-square statistics with imputed data, Unpublished Ph.D. dissertation, The University of Michigan, Department of Biostatistics (1984).

[5] P.A. Gimotty and M.B. Brown, Goodness-of-fit chi-square tests with imputed data, in: *Proceedings of the Social Statistics Section, American Statistical Association* (1984) 292–295.

[6] B.L. Ford, An overview of hot-deck procedures, in: W.G. Madow, I. Oklin, and D.B. Rubin (Eds.), *Incomplete Data and Sample Surveys, Volume 2: Theory and Bibliographies* (Academic Press, New York, 1983).

[7] R.A. Kronmal and A.V. Peterson, On the alias method for generating random variables from a discrete distribution, *Amer. Statist.* **33** (1979) 214–218.

[8] H. Nisselson, Overview, in: W.G. Madow, H. Nisselson, and I. Olkin (Eds.), *Incomplete Data in Sample Surveys, Volume 1: Report and Case Studies* (Academic Press, New York, 1983).

[9] D.R. Rubin, Inference and missing data, *Biometrika* **63** (1976) 581–592.

[10] J. Sheil and I. O'Muircheartaigh, The distribution of nonnegative quadratic forms in normal variables, *Appl. Statist.* **26** (1977) 92–98.