Proceedings of Statistics Canada Symposium 2002
Modelling Survey Data for Social and Economic Research

WESVAR: SOFTWARE FOR COMPLEX SURVEY DATA ANALYSIS

G. Hussain Choudhry and Richard Valliant[1]

ABSTRACT

Nearly all surveys use complex sample designs to collect data and these data are frequently used for statistical analyses beyond the estimation of simple descriptive parameters of the target population. Many procedures available in popular statistical software packages are not appropriate for this purpose because the analyses are based on the assumption that the sample has been drawn with simple random sampling. Therefore, the results of the analyses conducted using these software packages would not be valid when the sample design incorporates multistage sampling, stratification, or clustering. We discuss WesVar software that computes estimates and replication variance estimates by properly reflecting complex sampling and estimation procedures. We also illustrate the WesVar features by using data from two Westat surveys that employ complex sample designs: the Third International Mathematics and Science Study (TIMSS), and the National Health and Nutrition Examination Survey (NHANES).

Key Words: Weighting Survey Data, Nonresponse Adjustment, Poststratification, Raking, Two-way and Multi-way Tables, Quantile Estimation, Logit and Multinomial Models, Replication Variance Estimation.

1. INTRODUCTION

Use of standard statistical techniques is not appropriate for analyzing data collected in complex surveys. Specialized software is usually required to account for features like selection with varying probabilities and nonindependent selections. Estimates from complex survey data, like ratio means, regression coefficients, and odds ratios are themselves complicated, and methods of standard error estimation are needed that account for these complexities.

WesVar computes estimates and replication variance estimates that do properly reflect complex sampling and estimation procedures. Replication variance estimation consists of repeatedly calculating estimates for subgroups of the full sample and then computing the variance among these "replicate" estimates.

WesVar is flexible and can be used with a wide range of complex sample designs, including multistage, stratified, and unequal probability samples. The replicate variance estimates can also reflect many types of estimation schemes, such as nonresponse adjustment and ratio estimation (e.g., poststratification and raking). WesVar's powerful features and user-friendly Windows® interface make it easy to create replicate weights and use them for analysis or to import and analyze files that already contain replicate weights.

WesVar can calculate estimates of statistics such as totals and means, along with their standard error estimates. It is also easy to use WesVar to compute variance estimates for complex functions of estimates, e.g., ratios, differences of ratios, and log-odds ratios, based on tabular data. WesVar can also estimate coefficients for linear and logistic regression models and test the significance of linear combinations of parameter estimates.

Section 2 of this paper gives an overview of replication methods that are included in WesVar. The different types of weights that can be calculated are discussed in Section 3, and estimates in tables and regression models are described in Sections 4 and 5. Examples of analyses conducted with WesVar using data from complex surveys are included in section 6. Version 4.2 of WesVar has recently been released to the public. More detailed descriptions of its features are available at www.westat.com/wesvar. One of the most useful enhancements in WesVar version 4.2 is its ability to directly import files from a variety of formats, including SAS® (sd2, sas7bdat, ssd, transport), SPSS®, Stata®, and Microsoft Excel® and Access®.

[1]Westat, Inc. 1650 Research Boulevard, Rockville, Maryland 20850, USA.

2. OVERVIEW OF REPLICATION METHODS

The basic idea behind replication is to select subsamples repeatedly from the whole sample, calculate the statistic of interest for each subsample, and then use these subsample or replicate statistics to estimate the variance of the full-sample statistic. Different ways of creating subsamples from the full sample result in different replication methods. The subsamples are called *replicates* and the statistics calculated from these replicates are called *replicate estimates.* WesVar supports the following replication methods of variance estimation:

- Balanced Repeated Replication for stratified designs with two primary sampling units (PSUs) per stratum (BRR);
- Fay's BRR variant (FAY);
- Jackknife for unstratified designs (JK1);
- Jackknife for stratified designs with two PSUs per stratum (JK2); and
- Jackknife for stratified designs with two or more PSUs per stratum (JKn).

Other methods of replication such as the bootstrap can be handled in WesVar, but you must input the replicate weights and factors appropriate for that method.

Suppose that $\hat{\theta}$ is the full-sample estimate of some population parameter θ. The replication variance estimator, $v\left(\hat{\theta}\right)$, computed by WesVar takes the form

$$v\left(\hat{\theta}\right) = c\sum_{g=1}^{G} f_g h_g \left(\hat{\theta}_{(g)} - \hat{\theta}\right)^2 , \tag{2.1}$$

where

$\hat{\theta}_{(g)}$ is the estimate of θ based on the observations included in the *g*-th replicate;

G is the total number of replicates formed; and

c is a constant that depends on the replication method.

The factor f_g is a finite population correction that can be used with the jackknife methods; h_g is a scaling factor used only for the jackknife methods.

One of the main advantages of replication is its ease of use at the analysis stage. The same estimation procedure is used for the full sample and for each replicate. The variance estimates are then readily computed by a simple procedure. Furthermore, the same procedure is applicable to most statistics including means, percentages, ratios, regression coefficients, and combinations like differences. These estimates can also be calculated for analytic groups or subpopulations. A user need not understand the sampling or estimation methods if replicate weights are included with the data.

Another important advantage of replication is that it provides a simple way to account for adjustments that are made in weighting. Frequently, sampling weights are adjusted for nonresponse, poststratification or raking to control totals. By separately computing the weighting adjustments for each replicate, it is possible to reflect the effects of weight adjustments in the estimates of variance. By doing so, replication variance estimates have desirable design-based and model-based statistical properties. Shao (1996) reviews the methods and their design-based properties in finite population estimation while Valliant, Dorfman, and Royall (2000) cover model-based properties. Appendix D of the WesVar manual (Westat 2000) gives a detailed discussion of how to construct replicates for some common sample designs. That appendix also contains a flowchart of the issues to consider when forming replicates.

2.1 Balanced Repeated Replication (BRR) and Fay's Method

Balanced repeated replication applies to single-stage or multi-stage designs where the population of PSUs can be grouped into *L* variance strata (referred to as VarStrat in WesVar), with two PSUs (referred to as VarUnits) selected

from each stratum. For designs that do not fit this standard form, strata or PSUs can often be legitimately grouped, as discussed in Appendix D of the WesVar manual, to create a two-per-stratum design.
Each replicate half-sample estimate is formed by selecting one of the two VarUnits from each VarStrat based on a Hadamard matrix (see McCarthy 1969). Then, only the selected VarUnits are used to estimate the parameter of interest. Hadamard matrices of various sizes (up to 512) are stored and the same matrix is applied for each file that has the same number of VarStrat. These matrices give orthogonally balanced sets of replicates (see Wolter 1985, p. 115). WesVar will create more than 512 replicate weights for BRR (or Fay's method) if you supply an appropriate Hadamard matrix in a text file. The maximum matrix size is 9,984 by 9,984.

The Fay's method (Fay 1989) is a variant of BRR that has better properties in certain situations. Standard BRR can run into problems when computing an estimate for a small domain or estimating a ratio if the denominator has few sample cases. Fay's method corrects this problem by retaining all sample units in each replicate while modifying the sample weights differently than in the standard BRR.

2.2 Jackknife Methods

The Jackknife 1 (JK1) is appropriate when explicit stratification has not been used to select the sample. To form the replicates for JK1, identify G subsets using VarUnit. If a subset (i.e., a VarUnit) is a single sample unit, then JK1 is the standard delete-one jackknife. Replicates are formed by deleting one VarUnit at a time and multiplying the weights for the other VarUnits by $G/(G-1)$ where G is the number of replicates. When each subset to be deleted is a randomly formed group of units, the JK1 method is essentially the same as the nonoverlapping random group method discussed in Wolter (1985, Chapter 2). The maximum number of jackknife replicates is 9,999.

The basic sample design assumed for the Jackknife 2 (JK2) method is the same as that used for BRR — two PSUs (VarUnits) are sampled in each of L strata (VarStrat). In the case of a two-per-stratum design, there is a simplification of the jackknife that occurs for linear estimators that is used for JK2. The JKn method is more general and can be used when the number of PSUs (VarUnits) in a stratum (VarStrat) is greater than or equal to 2. The number of replicates, G, is equal to $\sum_{h=1}^{L} n_h$ where L is the number of VarStrat and n_h is the number of VarUnits in VarStrat h.

2.3 Degrees of Freedom for Variance Estimates

The lower and upper bounds of confidence intervals and the p value for test statistics are based on a t statistic with degrees of freedom (DF) determined by the method of variance estimation. Rust and Rao (1996) give theory for the DF approximations. An alternative is to assume an infinite number of degrees of freedom in which case the normal approximation is used. For the variance estimation methods offered by WesVar, the default numbers of degrees of freedom are based on the number of VarStrat and replicates and are discussed in Appendix A of the WesVar manual.

2.4 Finite Population Correction Factors

The theory for replication methods assumes that the first stage sampling units have been selected with replacement, or if not, that the design can be safely treated as if with replacement sampling had been used. There is a limited capability to introduce a finite population correction (FPC) for the jackknife methods but not for BRR or Fay's method. The FPCs must be associated with individual replicates, as shown in expression (2.1), and are discussed in detail in Appendix A of the manual.

The FPC factors can be in a separate file or can be specified on the *Attach Factors* screen.

3. WEIGHTING WITH WESVAR

The first step in weighting is to define the method of variance estimation to be used (BRR, Fay, JK1, JK2, or JKn). The user specifies the variance strata (VarStrat) and the PSUs (VarUnit) within each stratum, and the full sample base weights. The software then calculates replicate base weights appropriate for the selected method of variance estimation. WesVar also allows nonresponse adjustment, and poststratification and raking adjustments.

3.1 Nonresponse Adjustment

WesVar calculates nonresponse adjustments using the method of weighting classes. Both eligible responding and nonresponding units are classified into cells and the nonresponse adjustment is calculated within each cell. WesVar computes the nonresponse adjustments for both the full sample and the replicate weights. Both full sample weights and replicate weights need to be on the data file prior to nonresponse adjustment. The outputs are the full sample and replicate nonresponse-adjusted weights on a newly created WesVar data file that has the same number of records as the input data file. Nonrespondents will be on the file with weights of zero. Any ineligible units will still be on the file with their original weights. These ineligible cases may need to be eliminated using the subpopulation or subset features when doing tabulations.

By using the nonresponse adjustment procedure, WesVar can also adjust sample weights for unknown eligibility. Adjustment factors for this procedure are calculated in the same fashion as nonresponse adjustments. If this adjustment is used, it must be done before the nonresponse adjustment. The adjustment for unknown eligibility is required when the eligibility of all the sample cases cannot be determined. In the first step the base weights of sample cases with unknown eligibility are distributed proportionally over those with known eligibility. In the second step, the eligibility-adjusted weights of nonrespondents are proportionally distributed over the respondents. The weighting classes for unknown eligibility and nonresponse adjustments should be determined separately for each adjustment.

3.2 Poststratification

To use poststratification, the user must provide a file with poststrata cell identifiers and control totals. Note that the cell identifier can be constructed from a combination of other variables so that poststratification is not necessarily limited to a single variable. WesVar computes the poststratification adjustments for both the full sample and the replicate weights. If the file already contains poststratified full sample and replicate weights, then these weights should be used in analyzing the data. No special statements are required to notify the program that the weights being input are poststratified.

3.3 Raking

To use WesVar's Raking function, specify a text file that contains the control totals for each dimension. The fields in this text file should include the level of the variable and the corresponding control total. WesVar allows for raking with up to a maximum of eight dimensions. Raking is an iterative procedure. The user can specify the maximum number of iterations or tolerances on the absolute or relative distance that marginal estimates are from the controls. Processing stops when the specified number of iterations is completed or one of the user-specified stopping rules is satisfied. The default number of iterations is four; the maximum number of iterations is 100.

3.4 Self-Representing Units

Wesvar can handle designs with self-representing (SR) units. Note that in multi-stage designs, units that are contained in a SR unit at one stage but are sampled at another are not SR units. This procedure should only be used for units that are absolute certainties, and WesVar displays a message that warns the user that the identified SR units are treated as absolute certainties in the calculation of variances.

The specification of SR units depends on the variance estimation method. If JK1 is used, then the variable that defines VarUnit is used to identify the SR units. For the other methods, the VarStrat variable is used to identify the SR units. Once the SR units are identified, the replicate weights that are associated with non-SR units are created. The number of replicates depends on the number of non-SR units for JK1, or the number of non-SR strata for other methods.

4. ESTIMATES FROM TABLES REQUEST

Creating estimates and their standard errors in tables is largely controlled in WesVar by specifying *Table Request* options such as *Analysis Variables, Table Variables, Computed Statistics* and *Cell Function Statistics.* A *Table Request* allows you to analyze complex survey data by producing statistics such as totals, ratio means, proportions, general ratios, or other functions of totals. The *Analysis Variables* option in a *Table Request* allows you to specify the numeric variables for which population aggregates are to be estimated (such as income). The *Computed Statistics* option is used to create estimates that are functions of estimated totals.

Frequently, statistics are needed for subgroups (or domains) of the population and the analysis often requires the use of crosstabulations. The *Table Set* option of a *Table Request* can be used to specify the subgroups defined by a single categorical variable, as well as subgroups defined by crosstabulating two or more categorical variables. Within a table, *Cell Function Statistics* are used to create estimates that are functions of the estimates in two or more cells of a table.

4.1 Estimates of Totals and Ratios

A *Table Request* operates by calculating weighted totals for the specified variables of interest. The estimated general ratio is the ratio of two estimated totals. The ratio means and proportions are special cases of the general ratio estimates.

4.2 Medians and Quantiles

The median and the value at any whole percentile point can also be estimated in a *Table Request*. The use of replication methods for the direct estimation of variances of estimated percentiles, particularly the median, has been and continues to be an active area of research. The research by Kovar, Rao, and Wu (1988) indicates that the Jackknife method performs poorly for estimates of quantiles, whereas BRR and Fay's methods work well (Rao and Shao 1999). WesVar can compute the variances of quantiles indirectly using the Woodruff method (see Särndal, Swensson, and Wretman 1992), or directly by replication.

The methods for computing variances for quantiles are the Group and No Group methods. The Group method is included as a means of limiting computations on large data sets. The number of groups can be a value from 3 to 500 (the default is 50).

4.3 Computed Statistics

A number of functions are available on the *Computed Statistics* panel, e.g., Mean, GeoMean, Median, Quantiles, Log, etc. Other more complex *Computed Statistics* can also be specified in WesVar. Note that if a table variable is specified, the expression is evaluated for each crossclassification of the table variables.

4.4 Plausible Values or Multiple Imputations

The theory of plausible value (PV) estimation in education achievement assessments is due to Mislevy and Sheehan (1989). Their work is based on more general procedure of multiple imputation described by Rubin (1987). Suppose that we have *M* plausible values and the estimates of a parameter θ from these PVs are $\hat{\theta}_1$, $\hat{\theta}_2$, … $\hat{\theta}_M$. Then, the

algorithm for combining the results of individual repeated analyses to estimate the parameter θ and its variance is as follows.

The estimator of the parameter is the average of the PV estimates, i.e., $\hat{\theta}^* = M^{-1}\sum_{m=1}^{M}\hat{\theta}_m$. The variance of $\hat{\theta}^*$ is computed using formulas specific to PVs or multiple imputation. If we denote the replication variance of $\hat{\theta}_m$ as v_m, then the final estimate of the variance is calculated as

$$v\left(\hat{\theta}^*\right) = \frac{1}{M}\sum_{m=1}^{M} v_m + \left(1+\frac{1}{M}\right)\times\frac{1}{M-1}\sum_{m=1}^{M}\left(\hat{\theta}_m - \hat{\theta}^*\right)^2,$$

where the first term is the "within" variance component and the second term is the "between" variance component.

4.5 Standardised Rates

WesVar can calculate standardised rates using the direct standardisation method. Rates are adjusted using control totals (or standard distribution) so that the effects of population composition are eliminated when making comparisons between groups. For example, death rates in two countries may be standardized by age, so that comparisons of national death rates are not affected by differences in the age distributions of the countries.

4.6 Differences and other Complex Estimates

WesVar can also be used to perform analyses involving complex functions of estimated totals by defining the *Computed Statistics*, specifying table variables, and defining a function of the cell estimates on the *Cell Functions* screen. The table variable defines the cells in the crosstabulation. It is important to make the distinction between the *Computed Statistics* and *Cell Function Statistics*. The *Cell Function Statistics* allows differences (or more complex functions) between the cells of a crosstabulation to be calculated for the same variable in different subpopulations, while the *Computed Statistics* is used to calculate the differences across the entire population between different variables.

4.7 Design Effects

The design effect (DEFF) computed by WesVar is the ratio of the variance under the actual survey design to the variance under simple random sampling with replacement (SRSWR). This definition of DEFF differs from that of Kish (1965), who uses the variance from simple random sampling without replacement in the denominator. Moreover, the SRSWR variance is conditional on the achieved sample size for the domain of interest. For multi-way tables though, the DEFF is computed based on the sample size for the two-way marginal.

4.8 Confidence Intervals for Proportions

The default method of constructing a confidence interval for a percentage uses a symmetric interval of the form $\hat{p} \pm t_\nu\sqrt{v(\hat{p})}$, where $\hat{p}$ is the estimated percentage, $v(\hat{p})$ is its estimated variance, and t_ν is a multiplier from the *t*-distribution with ν degrees of freedom. A defect of this method is that for extreme percentages the upper or lower confidence bounds can go beyond the acceptable range of [0,100]. In such situations, the Wilson score method (Newcombe 1998) can be used to calculate confidence intervals that will always remain within [0,100]. Unlike the *t* approximation intervals, the Wilson intervals will not be symmetric around the point estimate of the percentage.

4.9 Chi-Square Statistics

Testing for independence in a two-way table can be done simply with a *Table Request*. For this purpose, Pearson's chi-square statistic is calculated, as well as two chi-square statistics, denoted by RS2 and RS3, that have been modified to reflect the complex sample design. The modified chi-square statistics rely on adjusting the Pearson chi-square statistics using an estimated "design effect", as suggested by Rao and Scott (1981, 1984).

4.10 Missing Data Procedures

If the input data set contains more than one representation of missing data, all of these representations are converted to one missing value representation for WesVar and are treated as the same missing value in all procedures.

If data are missing, a *Table Request* will still produce estimates and their standard errors under most circumstances. When defining subgroups of a table, the default is to exclude from the output any statistics for subgroups defined by a missing value on one of the categorical table variables. If any analysis variable or any variable used to define a *Computed Statistics* is missing the default is to delete the entire record from the request. Thus, WesVar restricts the data to those records with no missing values for all of the analysis variables and all of the variables used to form *Computed Statistics*. Cases with missing values are also excluded from regression requests.

5. MODELS IN REGRESSION

Users can request linear, logistic, or multinomial regression models. WesVar's regression procedures estimate the parameters of a regression model and provide a variety of other statistics, including a test for the overall fit of the model and of individual parameters, measures of fit, odds ratios, and other statistics. Appendix C of the WesVar manual gives detailed discussions of the computational methods used to solve for parameter estimates and other model statistics. We give only a brief sketch of the techniques here.

5.1 Linear Regression Model

WesVar fits linear regression models of the form $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$ where $\mathbf{Y}$ is the $n \times 1$ column vector of sample observations, $\beta = \left(\beta_1, \ldots, \beta_p\right)$ is the $p \times 1$ column vector of regression coefficients, $\mathbf{X}$ is the $n \times p$ matrix of independent variables and $\mathbf{e}$ is the $n \times 1$ column vector of random errors. The i-th row of $\mathbf{X}$ is the vector of explanatory variables for unit i, $\mathbf{x}_i' = \left(x_{i1}, x_{i2}, \ldots, x_{ip}\right)$. The vector $\mathbf{x}_i$ can contain continuous or discrete variables. The weighted least squares estimate of the parameter vector is $\mathbf{b} = \left(\mathbf{X}'\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{W}\mathbf{Y}$ where $\mathbf{W}$ is the $n \times n$ diagonal matrix formed from the n full sample weights $w_1, w_2, \ldots, w_n$.

WesVar also calculates the weighted least squares estimate for each replicate subsample. These replicate estimates are then combined with a matrix formula, analogous to (2.1), to give a replication estimate of the covariance matrix of $\mathbf{b}$. Elements of this covariance estimate are then used to construct t-tests and confidence intervals for individual coefficients and customized tests on linear combinations of regression coefficients.

5.2 Logistic Regression Model

In a logistic regression model, with $\mathbf{x}_i$ defined as above for linear regression, the expected value of a dichotomous variable Y_i is assumed to be $p\left(\mathbf{x}_i\right) = \dfrac{exp\left(\mathbf{x}_i'\boldsymbol{\beta}\right)}{1 + exp\left(\mathbf{x}_i'\boldsymbol{\beta}\right)}$. WesVar uses a pseudo maximum likelihood approach for parameter estimation in which the value $\mathbf{b}$ is found that maximizes the weighted sample log-likelihood. This estimate

is known as the pseudo maximum likelihood estimate (MLE). WesVar solves for the pseudo-MLE using a modified version of the Newton-Raphson method. Computational checks are also included for both convergence and divergence of parameter estimates.

WesVar computes three measures of fit for logistic regression models that are based on comparing the log-likelihood for the fitted model to that for a model that includes only the intercept. The three measures are known in the literature as negative log-likelihood (or entropy), Cox-Snell likelihood ratio, and Estrella likelihood ratio. All three of the measures are printed by default. The three alternatives, along with several others, are reviewed in Mittlböck and Schemper (1996) and Estrella (1998).

By default WesVar computes odds ratios only for main effects that are not involved in interactions. A two-sided confidence interval is also calculated for each odds ratio. The logarithm of the odds (or log-odds) that the response is a 1 for sample unit *i* is $log\left[\frac{p(\mathbf{x}_i)}{1-p(\mathbf{x}_i)}\right] = \mathbf{x}_i'\boldsymbol{\beta}$, which is also referred to as the logit of $p(\mathbf{x}_i)$. When a variable x_{ik} is continuous, a parameter β_k is the logarithm of the odds ratio of a 1-unit change in x_{ik}, holding all other variables constant. The quantity $exp(\beta_k)$ is, thus, the odds ratio of a 1-unit change in x_{ik} for any unit *i*. A two-sided confidence interval for an odds ratio is found by putting a confidence interval on the parameter β_k and then transforming to the odds ratio scale. The standard error of the estimate of β_k is estimated using the replication method specified by the user.

In a model that includes interactions, calculation of a meaningful odds ratio is more complicated, but WesVar has a tool that allows a user to compute a customized odds ratio by combining estimates of the model parameters.

5.3 Multinomial Logistic Regression Model

A multinomial logistic regression model or generalized logit model is an extension of the logistic regression model and is described in detail in Agresti (1990). In multinomial logistic regression, the response variable *Y* can be a categorical response with *K* categories. As in dichotomous logistic regression, WesVar uses the pseudo MLE method for parameter estimation, both in the full sample and the replicates.

The same three measures of fit — entropy, Cox-Snell, and Estrella — are computed in multinomial logistic as in dichotomous logistic regression. Odds ratios and confidence intervals are also calculated in a similar way. For multinomial logistic regression a default odds ratio is the ratio of the odds of being in category *k* of the multinomial versus the reference category *K* given a 1-unit change in one of the explanatory variables (holding the other predictor variables constant). WesVar also provides a tool for computing customized odds ratios that are not printed by default.

6. EXAMPLES USING WESVAR

We illustrate two examples of data analysis from Westat surveys using WesVar. These are the Third International Mathematics and Science Study (TIMSS) and the National Health and Nutrition Examination Survey (NHANES).

6.1 Third International Mathematics and Science Study (TIMSS)

The basic sample design implemented for TIMSS is generally a two-stage stratified cluster sample design. The first stage consists of a stratified sample of schools, and the second stage consists of samples of classrooms from each eligible target grade in sampled schools. In some countries, a third stage was added, in which students were sampled within classrooms. Table 6.1 provides a comparison between unweighted SAS, weighted SAS and the WesVar results. Without special programming SAS PROC MEANS procedure can handle one PV at a time. The weighted SAS standard errors do not account for clustering and are much smaller than the WesVar standard errors using only

the first PV or, more appropriately, all five PVs in the multiple imputation formula. Note that the "between" variance component for the PV analysis is small as compared with "within" component in this case.

Table 6.1: Comparison between Unweighted SAS, Weighted SAS, and WesVar Results.

Country	Test	Unweighted SAS		Weighted SAS		WesVar			
		First PV		First PV		First PV		5 PVs	
		Mean	SE	Mean	SE	Mean	SE	Mean	SE
Netherlands	Science	551.7	1.3	543.5	1.3	543.5	6.6	544.8	7.0
	Math	544.4	1.2	537.9	1.3	537.9	6.8	539.9	7.2
Belgium	Science	547.1	0.9	534.0	1.0	534.0	2.7	534.9	2.9
	Math	573.9	0.9	556.5	1.1	556.5	3.1	558.0	3.5

Note: TIMSS uses JK2 method of replication. Number of replicates is 74.

6.2 National Health and Nutrition Examination Survey (NHANES)

The NHANES sample represents the total non-institutionalized civilian population in the 50 states and the District of Columbia in the U.S. A four-stage sample design is being used. To reduce the amount of travel, PSUs are defined to be individual counties or groupings of adjacent counties. The second stage consists of area segments made up of Census blocks or combinations of blocks. The third stage of sample selection consists of households and non-institutional group quarters. Sample persons within the households or group quarters are the fourth stage. The sample PSUs and area segments are selected with probability proportional to size (PPS). The sample is designed to produce approximately equal sample size per PSU.

To illustrate logistic regression, we used a subset of the full sample collected in 1994 to model the presence of asthma (ASTHMA) as a function of whether a person had smoked 100 or more cigarettes in his/her lifetime (CIG100[2]), a 4-level race-ethnicity variable (RACETHN[4]), presence or absence of hayfever (HAYFEVER[2]) and gender (SEX[2]). WesVar uses the [*n*] notation to denote a categorical variable with *n* levels. WesVar automatically creates dummy variables for each categorical variable and sets the parameter solution for the highest level of each to zero to obtain a set of solutions.

Variances were estimated using Fay's method with 24 replicates. By default, WesVar prints parameter estimates and various hypotheses tests, which we omit because of space limitations. The F-value for overall fit was 30.35 with 6 and 18 degrees of freedom, which is highly significant. Table 6.2 shows odds ratios for the main effects. For example, CIG100.1 denotes the first level of CIG100, i.e., the person has smoked 100 or more cigarettes. The odds ratio of having asthma for someone who has smoked this many cigarettes compared to someone who has not is 1.65 with a 95% confidence interval of [1.27, 2.13]. We also present a user-constructed odds ratio for comparing female smokers with hayfever to male nonsmokers who do not have hayfever (irrespective of race-ethnicity). On the logit scale the difference in the expected values for the two groups is CIG100.1 + HAYFEVER.1-SEX.1. The odds ratio is denoted by OR1 in Table 6.2 and is equal to 9.63 with a 95 % confidence interval of [5.52, 16.81].

Table 6.2 Odds ratio results

Parameter	Estimate	Lower 95%	Upper 95%
CIG100.1	1.65	1.27	2.13
RACETHN.1	0.72	0.47	1.11
RACETHN.2	0.90	0.59	1.38
RACETHN.3	0.53	0.37	0.77
HAYFEVER.1	4.45	3.24	6.11
SEX.1	0.76	0.59	0.98
OR1	9.63	5.52	16.81

REFERENCES

Agresti, A. (1990). *Categorical Data Analysis*. New York: John Wiley & Sons.

Estrella, A. (1998). A New Measure of Fit for Equations with Dichotomous Dependent Variables. *Journal of Business and Economic Statistics*, **16**, pp. 198-205.

Fay, R.E. (1989). Theoretical Application of Weighting for Variance Calculation. *Proceedings of the Section on Survey Research Methods of the American Statistical Association*, pp. 212-217.

Kish, L. (1965). *Survey Sampling*. New York: John Wiley & Sons.

Kovar, J.G., Rao, J.N.K., and Wu, C.F.J. (1988). Bootstrap and Other Methods to Measure Errors in Survey Estimates. *Canadian Journal of Statistics*, **16**, pp. 25-45.

McCarthy, P.J. (1969). Pseudo-replication: Half-samples. *Review of the International Statistical Institute*, **37**, pp. 239-264.

Mislevy, R.J., and Sheehan, K.M. (1989). The Role of Collateral Information about Examinees in Item Parameter Estimation. *Psychometrika*, **54**, pp. 661-679.

Mittlböck, M., and Schemper, M. (1996). Explained Variation for Logistic Regression. *Statistics in Medicine*, **15**, pp. 1987-1997.

Newcombe, R.G. (1998). Two-Sided Confidence Intervals for the Single Proportion: Comparison of Seven Methods. *Statistics in Medicine*, **17**, pp. 857-872.

Rao, J.N.K., and Scott, A.J. (1981). The Analysis of Categorical Data from Complex Sample Surveys: Chi-Squared Tests for Goodness of Fit and Independence in Two-Way Tables. *Journal of the American Statistical Association*, **76**, pp. 221-230.

Rao, J.N.K., and Scott, A.J. (1984). On Chi-Squared Tests for Multiway Contingency Tables with Cell Proportions Estimated from Survey Data. *The Annals of Statistics*, **12**, pp. 46-60.

Rao, J.N.K., and Shao, J. (1999). Modified Balanced Repeated Replication for Complex Survey Data. *Biometrika*, **86**, pp. 403-415.

Rubin, D. (1987). *Multiple Imputations for Nonresponse in Sample Surveys*. New York: John Wiley & Sons.

Rust, K., and Rao, J.N.K. (1996). Variance estimation for complex estimators in sample surveys. *Statistics in Medical Research*, **5**, pp. 381-397.

Särndal, C.E., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Shao, J. (1996). Resampling Methods in Sample Surveys, (with Discussion). *Statistics*, **27**, pp. 203-254.

Valliant, R., Dorfman, A.H., and Royall, R.M. (2000). *Finite Population Sampling and Inference: A Prediction Approach*. New York: John Wiley & Sons.

Westat (2000). *WesVar 4.0 User's Guide*. Rockville MD: Westat.

Wolter, K.M. (1985). *Introduction to Variance Estimation*. New York: Springer-Verlag.