(1999, chap. 4) discuss various issues dealing with weighted and unweighted estimates of population parameters and offer a measure of the inefficiency of using weighted estimates. They recommend using the weighted analysis if the inefficiency is not unacceptably large, to avoid the bias in the unweighted analysis. If the inefficiency is unacceptably large, they recommend using the unweighted analysis, augmenting the model with survey design variables, including the weight, to reduce the bias. Incorporation of design variables into the model is often problematic, however, because the inclusion of the design variables as additional covariates in the model may contradict the scientific purpose of the analysis. For example, when the objective of the analysis is to examine associations between health measures and risk factors, conditioning on the design variables may interfere with the relational pathway.

In complex large-scale surveys, it is often not possible to include in the model all the design information, especially when the sample weights are modified for nonresponse and poststratification adjustments (Alexander, 1987). Another practical problem in incorporating the design variables into the model is the lack of relevant information in the data set. Not all design-related variables are available to the analysts. Most public-use survey data provide only PSU (the primary sampling unit), leaving out secondary cluster units (census enumeration districts or telephone exchanges). Often, provision of secondary units may not be possible because of confidentiality issues.

In a model-based analysis, one must guard against possible misspecification of the model and possible omission of covariates. The use of sample weights (design-based analysis) can provide protection against model misspecification (DuMouchel & Duncan, 1983; Pfeffermann & Homes, 1985). Kott (1991) points out that the sampling weights need to be used in linear regression because the choice of covariates in survey data is limited in most secondary analyses. The merits and demerits of using sample weights will be further discussed in the last section of Chapter 6.

4. STRATEGIES FOR VARIANCE ESTIMATION

The estimation of the variance of a survey statistic is complicated not only by the complexity of the sample design, as seen in the previous chapters, but also by the form of the statistic. Even with an SRS design, the variance estimation of some statistics requires nonstandard estimating techniques. For example, the variance of the median is conspicuously absent in the standard texts, and the sampling error of a ratio estimator (refer again to Note 1) is complicated because both the numerator and denominator are random variables.

Certain variance estimating techniques not found in standard textbooks are sufficiently flexible to accommodate both the complexities of the sample design and the various forms of statistics. These general techniques for variance estimation, to be reviewed in this chapter, include replicated sampling, balanced repeated replication (BRR), jackknife-repeated replication (JRR), the bootstrap method, and the Taylor series method.

Replicated Sampling: A General Approach

The essence of this strategy is to facilitate the variance calculation by selecting a set of replicated subsamples instead of a single sample. It requires that each subsample be drawn independently and use of an identical sample selection design. Then an estimate is made in each subsample by the identical process, and the sampling variance of the overall estimate (based on all subsamples) can be estimated from the variability of these independent subsample estimates. This is the same idea as the repeated systematic sampling mentioned in Chapter 2.

The sampling variance of the mean ($\bar{u}$) of t replicate estimates u_1, $u_2, \ldots, u_t$ of the parameter U can be estimated by the following simple variance estimator (Kalton, 1983, p. 51):

$$v(\bar{u}) = \sum (u_i - \bar{u})^2 / t(t-1) \qquad (4.1)$$

This estimator can be applied to any sample statistic obtained from independent replicates of any sample design.

In applying this variance estimator, 10 replicates are recommended by Deming (1960), and a minimum of 4 by others (Sudman, 1976) for descriptive statistics. An approximate estimate of standard error can be calculated by dividing the range in the replicate estimates by the number of replicates when the number of replicates is between 3 and 13 (Kish, 1965, p. 620). However, because this variance estimator with t replicates is based on $(t-1)$ degrees of freedom for statistical inference, a larger number of replicates may be needed for analytic studies, perhaps 20 to 30 (Kalton, 1983, p. 52).

To understand the replicated design strategy, let us consider a simple example. Suppose we want to estimate the proportion of boys among 200 newly born babies. We will simulate this survey using the random digits from Cochran's book (1977, p. 19), assuming the odd numbers represent boys. The sample is selected in 10 replicate samples of $n = 20$ from the first 10 columns of the table. The numbers of boys in the replicates are as follows:

Replicate:	*9*	*8*	*13*	*12*	*14*	*8*	*10*	*7*	*10*	*8*	*Total = 99*
Proportion of Boys:	.45	.40	.65	.60	.70	.40	.50	.35	.50	.40	Proportion = .495

The overall percentage of boys is 49.5%, and its standard error is 3.54% $(= \sqrt{49.5*50.5/200})$. The standard error estimated from the 10 replicate estimates using Equation 4.1 is 3.58%. It is easy to get an approximate estimate of 3.50% by taking one tenth of the range (70%–35%). The chief advantage of replication is ease in estimation of the standard errors.

In practice, the fundamental principle of selecting independent replicates is somewhat relaxed. For one thing, replicates are selected using sampling without replacement instead of with replacement. For unequal probability designs, the calculation of basic weights and the adjustment for nonresponse and poststratification usually are performed only once for the full sample, rather than separately within each replicate. In cluster sampling, the replicates often are formed by systematically assigning the clusters to the t replicates in the same order that the clusters were first selected, to take advantage of stratification effects. In applying Equation 4.1, the sample mean from the full sample generally is used for the mean of the replicate means. These deviations from fundamental principles can affect the variance estimation, but the bias is thought to be insignificant in large-scale surveys (Wolter, 1985, pp. 83–85).

The community mental health survey conducted in New Haven, Connecticut, in 1984 as part of the ECA Survey of the National Institute of Mental Health (E. S. Lee, Forthofer, Holzer, & Taube, 1986) provides an example of replicated sampling. The sampling frame for this survey was a geographically ordered list of residential electric hookups. A systematic sample was drawn by taking two housing units as a cluster, with an interval of 61 houses, using a starting point chosen at random. A string of clusters in the sample was then sequentially allocated to 12 subsamples. These subsamples were created to facilitate the scheduling and interim analysis of data during a long period of screening and interviewing. Ten of the subsamples were used for the community survey, with the remaining two reserved for another study. The 10 replicates are used to illustrate the variance estimation procedure.

These subsamples did not strictly adhere to a fundamental principle of independent replicated sampling because the starting points were systematically selected, except for the first random starting point. However, the systematic allocation of clusters to subsamples in this case introduced an approximate stratification leading to more stable variance estimation and,

TABLE 4.1
Estimation of Standard Errors From Replicates:
ECA Survey in New Haven, 1984 ($n = 3{,}058$)

			Regression Coefficients[a]			
Replicate	*Prevalence Rate*[b]	*Odds Ratio*[c]	*Intercept*	*Gender*	*Color*	*Age*
Full Sample	17.17	0.990	0.2237	−0.0081	0.0185	−0.0020
1	12.81	0.826	0.2114	0.0228	0.0155	−0.0020
2	17.37	0.844	0.2581	0.0220	0.0113	−0.0027
3	17.87	1.057	0.2426	−0.0005	0.0393	−0.0015
4	17.64	0.638	0.1894	0.0600	0.2842	−0.0029
5	16.65	0.728	0.1499	0.0448	−0.0242	−0.0012
6	18.17	1.027	0.2078	−0.0024	−0.0030	−0.0005
7	14.69	1.598	0.3528	−0.0487	−0.0860	−0.0028
8	17.93	1.300	0.3736	−0.0333	−0.0629	−0.0032
9	17.86	0.923	0.2328	−0.0038	0.0751	−0.0015
10	18.91	1.111	0.3008	−0.0007	0.0660	−0.0043
Range	6.10	0.960	0.2237	0.1087	0.3702	0.0038
Standard error based on:						
Replicates	0.59	0.090	0.0234	0.0104	0.0324	0.0004
SRS	0.68	0.097	0.0228	0.0141	0.0263	0.0004

SOURCE: Adapted from "Complex Survey Data Analysis: Estimation of Standard Errors Using Pseudo-Strata," E. S. Lee, Forthofer, Holzer, and Taube, *Journal of Economic and Social Measurement*, © *copyright* 1986 by the *Journal of Economic and Social Measurement.* Adapted with permission.

a. The dependent variable (coded as 1 = condition present and 0 = condition absent) is regressed on sex (1 = male, 0 = female), color (1 = black, 0 = nonblack), and age (continuous variable). This analysis is used for demonstration only.

b. Percentage with any mental disorders during the last 6 months.

c. Sex difference in the 6-month prevalence rate.

therefore, may be preferable to a random selection of a starting point for this relatively small number of replicates. Therefore, we considered these subsamples as replicates and applied the variance estimator with replicated sampling, Equation 4.1.

Because one adult was randomly selected from each sampled household using the Kish selection table (Kish, 1949), the number of adults in each household became the sample case weight for each observation. This weight was then adjusted for nonresponse and poststratification. Sample weights were developed for the full sample, not separately within each subsample, and these were the weights used in the analysis.

Table 4.1 shows three types of statistics calculated for the full sample as well as for each of the replicates. The estimated variance of the prevalence rate, in percent (p), can be calculated from the replicate estimates (p_i) using Equation 4.1:

$$v(p) = \frac{\sum (p_i - 17.17)^2}{10(10-1)} = 0.3474,$$

and the standard error is $\sqrt{0.3474} = 0.59$. The overall prevalence rate of 17.17% is slightly different from the mean of the 10 replicate estimates because of the differences in response rates. Note that one tenth of the range in the replicate estimates (0.61) approximates the standard error obtained by Equation 4.1. Similarly, standard errors can be estimated for the odds ratio and regression coefficients. The estimated standard errors have approximately the same values as those calculated by assuming simple random sampling (using appropriate formulas from textbooks). This indicates that design effects are fairly small for these statistics from this survey.

Although the replicated sampling design provides a variance estimator that is simple to calculate, it requires a sufficient number of replicates to obtain acceptable precision for statistical inference. But if there is a large number of replicates and each replicate is relatively small, it severely limits using stratification in each replicate. Most important, it is impractical to implement replicated sampling in complex sample designs. For these reasons, a replicated design is seldom used in large-scale, analytic surveys. Instead, the replicated sampling idea has been applied to estimate variance in the data analysis stage. This attempt gave rise to pseudo-replication methods for variance estimation. The next two techniques are based on this idea of pseudo-replication.

Balanced Repeated Replication

The balanced repeated replication (BRR) method is based on the application of the replicated sampling idea to a paired selection design in which two PSUs are sampled from each stratum. The paired selection design represents the maximum use of stratification, yet allows the calculation of variance. In this case, the variance between two units is one half of the squared difference between them. To apply the replicated sampling idea, we first divide the sample into random groups to form pseudo-replicates. If it is a stratified design, it requires all the strata to be represented in each pseudo-replicate. In a stratified, paired selection design, we can form only two pseudo-replicates: one containing one of the two units from each stratum and the other containing the remaining unit from each stratum (complement replicate). Each pseudo-replicate then includes approximately half of the total sample. Applying Equation 4.1 with $t = 2$, we can estimate the sampling variance of the mean of the two replicate estimates, u', u'', by

$$v(\bar{u}) = [(u' - \bar{u})^2 + (u'' - \bar{u})^2]/2. \tag{4.2}$$

As seen above, the mean of replicate estimates is often replaced by an overall estimate obtained from the full sample. However, this estimator is too unstable to have any practical value because it is based on only two pseudo-replicates. The BRR method solves this problem by repeating the process of forming half-sample replicates, selecting different units from different strata. The pseudo-replicated half samples then contain some common units, and this introduces dependence between replicates, which complicates the estimation. One solution, which leads to unbiased estimates of variance for linear statistics, is to balance the formation of pseudo-replicates by using an orthogonal matrix (Plackett & Burman, 1946). The full balancing requires that the size of the matrix be a multiple of four and the number of replicates be greater than or equal to the number of strata. Then the sampling variance of a sample statistic can be estimated by taking the average of variance estimates by Equation 4.2 over t pseudo-replicates:

$$v(\bar{u}) = \sum [(u_i' - \bar{u})^2 + (u_i'' - \bar{u})^2]/2t = \sum (u_i'' - u_i'')^2/4t. \tag{4.3}$$

It is possible to reduce computation by dropping the complement half-sample replicates:

$$v'(\bar{u}) = \sum (u_i' - \bar{u})^2/t. \tag{4.4}$$

This is the estimator originally proposed by McCarthy (1966). This balancing was shown by McCarthy to yield unbiased estimates of variance for linear estimators. For nonlinear estimators, there is a bias in the estimates of variance, but numerical studies suggest that it is small. For a large number of strata, the computation can be further simplified by using a smaller set of partially balanced replicates (K. H. Lee, 1972; Wolter, 1985, pp. 125–130).

As in replicated sampling, BRR assumes that the PSUs are sampled with replacement within strata, although in practice sampling without replacement generally is used. Theoretically, this leads to an overestimation of variance when applied to a sample selected without replacement, but the overestimation is negligible in practice because the chance of selecting the same unit more than once under sampling without replacement is low when the sampling fraction is small. The sampling fraction in a paired selection design (assumed in the BRR method) usually is small because only two PSUs are selected from each stratum.

When used with a multistage selection design, BRR usually is applied only to PSUs and disregards the subsampling within the PSUs. Such a practice is predicated on the fact that the sampling variance can be approximated adequately from the variation between PSU totals when the first-stage sampling fraction is small. This is known as the ultimate cluster approximation. As shown in Kalton (1983, chap. 5), the unbiased variance estimator for a simple two-stage selection design consists of a component from each of the two stages, but the term for the second-stage component is multiplied by the first-stage sampling fraction. Therefore, the second-stage contribution becomes negligible as the first-stage sampling fraction decreases. This shortcut procedure based only on PSUs is especially convenient in the preparation of complex data files for public use as well as in the analysis of such data, because detailed information on complex design features is not required except for the first-stage sampling.

If the BRR technique is to be applied to other than the paired selection designs, it is necessary to modify the data structure to conform to the technique. In many multistage surveys, stratification is carried out to a maximum and only one PSU is selected from each stratum. In such case, PSUs can be paired to form collapsed strata to apply the BRR method. This procedure generally leads to some overestimation of the variance because some of the between-strata variability is now included in the within-stratum calculation. The problem is not serious for the case of linear statistics if the collapsing is carried out judiciously; however, the collapsing generally is not recommended for estimating the variance of nonlinear statistics (see Wolter, 1985, p. 48). The Taylor series approximation method discussed later may be used for the nonlinear statistics. Although it is not used widely, there is a method of constructing orthogonal balancing for three PSUs per stratum (Gurney & Jewett, 1975).

Now let us apply the BRR technique to the 1984 GSS. As introduced in the previous chapter, it used a multistage selection design. The first-stage sampling consisted of selecting one PSU from each of 84 strata of counties or county groups. The first 16 strata were large metropolitan areas and designated as self-representing (or automatically included in the sample). To use the BRR technique, the 84 strata are collapsed into 42 pairs of pseudo-strata. Because the numbering of non–self-representing PSUs in the data file followed approximately the geographic ordering of strata, pairing was done sequentially, based on the PSU code. Thus, the 16 self-representing strata were collapsed into 8 pseudo-strata, and the remaining 68 non–self-representing strata into 34 pseudo-strata. This pairing of the self-representing strata, however, improperly includes variability among them. To exclude this and include only the variability within each of the self-representing strata, the combined observations within each self-representing pseudo-stratum were randomly grouped into two pseudo-PSUs.

To balance the half-sample replicates to be generated from the 42 pseudo-strata, an orthogonal matrix of order 44 (see Table 4.2) was used. This matrix is filled with zeros and ones. To match with the 42 strata, the first two columns were dropped (i.e., 44 rows for replicates and 42 columns for pseudo-strata). A zero indicates the inclusion of the first PSU from the strata, and a one denotes the inclusion of the second PSU. The rows are the replicates, and the columns represent the strata. For example, the first replicate contains the second PSU from each of the 42 pseudo-strata (because all the elements in the first row are ones). Using the rows of the orthogonal matrix, 44 replicates and 44 complement replicates were created.

To estimate the variance of a statistic from the full sample, we needed first to calculate the statistic of interest from each of the 44 replicates and complement replicates. In calculating the replicate estimates, the adjusted sample weights were used. Table 4.3 shows the estimates of the proportion of adults approving the "hitting" for the 44 replicates and their complement replicates. The overall proportion was 60.0%. The sampling variance of the overall proportion, estimated by Equation 4.3, is 0.000231. Comparing this with the sampling variance of the proportion under the SRS design [$pq/(n-1) = 0.000163$, ignoring FPC], we get the design effect of 1.42 (= 0.000231/0.000163). The design effect indicates that the variance of the estimated proportion from the GSS is 42% larger than the variance calculated from an SRS of the same size. The variance by Equation 4.4 also gives similar estimates.

In summary, the BRR technique uses a pseudo-replication procedure to estimate the sampling variance and is primarily designed for a paired selection design. It also can be applied to a complex survey, which selects one PSU per stratum by pairing strata, but the pairing must be performed judiciously, taking into account the actual sample selection procedure. In most applications of BRR in the available software packages, the sample weights of the observations in the selected PSUs for a replicate are doubled to make up for the half of PSUs not selected. There is also a variant of BRR, suggested by Fay (Judkins, 1990), in creating replicate weights, which uses $2-k$ or k times the original weight, depending on whether the PSU is selected or not selected based on the orthogonal matrix ($0 \le k < 1$). This will be illustrated further in the next chapter.

Jackknife Repeated Replication

The idea of jackknifing was introduced by Quenouille (1949) as a nonparametric procedure to estimate the bias, and later Tukey (1958) suggested how that same procedure could be used to estimate variance. Durbin (1959) first used this method in his pioneering work on ratio estimation. Later,

TABLE 4.2
Orthogonal Matrix of Order 44

Rows	*Columns (44)*
1	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
2	1 0 1 0 0 1 0 1 0 0 1 1 1 0 1 1 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 1 0 0 0 1 1 0 1 0 1 1 0
3	1 0 0 1 0 0 1 0 1 0 0 1 1 1 0 1 1 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 1 0 0 0 1 1 0 1 0 1 1
4	1 1 0 0 1 0 0 1 0 1 0 0 1 1 1 0 1 1 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 1 0 0 0 1 1 0 1 0 1
5	1 1 1 0 0 1 0 0 1 0 1 0 0 1 1 1 0 1 1 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 1 0 0 0 1 1 0 1 0
6	1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 1 1 0 1 1 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 1 0 0 0 1 1 0 1
7	1 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 1 1 0 1 1 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 1 0 0 0 1 1 0
8	1 0 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 1 1 0 1 1 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 1 0 0 0 1 1
9	1 1 0 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 1 1 0 1 1 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 1 0 0 0 1
10	1 1 1 0 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 1 1 0 1 1 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 1 0 0 0
11	1 0 1 1 0 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 1 1 0 1 1 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 1 0 0
12	1 0 0 1 1 0 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 1 1 0 1 1 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 1 0
13	1 0 0 0 1 1 0 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 1 1 0 1 1 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 1
14	1 1 0 0 0 1 1 0 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 1 1 0 1 1 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0
15	1 0 1 0 0 0 1 1 0 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 1 1 0 1 1 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0
16	1 0 0 1 0 0 0 1 1 0 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 1 1 0 1 1 1 1 1 0 0 0 1 0 1 1 1 0 0 0
17	1 0 0 0 1 0 0 0 1 1 0 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 1 1 0 1 1 1 1 1 0 0 0 1 0 1 1 1 0 0
18	1 0 0 0 0 1 0 0 0 1 1 0 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 1 1 0 1 1 1 1 1 0 0 0 1 0 1 1 1 0
19	1 0 0 0 0 0 1 0 0 0 1 1 0 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 1 1 0 1 1 1 1 1 0 0 0 1 0 1 1 1
20	1 1 0 0 0 0 0 1 0 0 0 1 1 0 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 1 1 0 1 1 1 1 1 0 0 0 1 0 1 1
21	1 1 1 0 0 0 0 0 1 0 0 0 1 1 0 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 1 1 0 1 1 1 1 1 0 0 0 1 0 1
22	1 1 1 1 0 0 0 0 0 1 0 0 0 1 1 0 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 1 1 0 1 1 1 1 1 0 0 0 1 0
23	1 0 1 1 1 0 0 0 0 0 1 0 0 0 1 1 0 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 1 1 0 1 1 1 1 1 0 0 0 1
24	1 1 0 1 1 1 0 0 0 0 0 1 0 0 0 1 1 0 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 1 1 0 1 1 1 1 1 0 0 0
25	1 0 1 0 1 1 1 0 0 0 0 0 1 0 0 0 1 1 0 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 1 1 0 1 1 1 1 1 0 0
26	1 0 0 1 0 1 1 1 0 0 0 0 0 1 0 0 0 1 1 0 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 1 1 0 1 1 1 1 1 0
27	1 0 0 0 1 0 1 1 1 0 0 0 0 0 1 0 0 0 1 1 0 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 1 1 0 1 1 1 1 1
28	1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 1 0 0 0 1 1 0 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 1 1 0 1 1 1 1
29	1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 1 0 0 0 1 1 0 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 1 1 0 1 1 1
30	1 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 1 0 0 0 1 1 0 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 1 1 0 1 1
31	1 1 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 1 0 0 0 1 1 0 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 1 1 0 1
32	1 1 1 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 1 0 0 0 1 1 0 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 1 1 0
33	1 0 1 1 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 1 0 0 0 1 1 0 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 1 1
34	1 1 0 1 1 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 1 0 0 0 1 1 0 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 1
35	1 1 1 0 1 1 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 1 0 0 0 1 1 0 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1
36	1 1 1 1 0 1 1 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 1 0 0 0 1 1 0 1 0 1 1 0 0 1 0 0 1 0 1 0 0
37	1 0 1 1 1 0 1 1 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 1 0 0 0 1 1 0 1 0 1 1 0 0 1 0 0 1 0 1 0
38	1 0 0 1 1 1 0 1 1 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 1 0 0 0 1 1 0 1 0 1 1 0 0 1 0 0 1 0 1
39	1 1 0 0 1 1 1 0 1 1 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 1 0 0 0 1 1 0 1 0 1 1 0 0 1 0 0 1 0
40	1 0 1 0 0 1 1 1 0 1 1 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 1 0 0 0 1 1 0 1 0 1 1 0 0 1 0 0 1
41	1 1 0 1 0 0 1 1 1 0 1 1 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 1 0 0 0 1 1 0 1 0 1 1 0 0 1 0 0
42	1 0 1 0 1 0 0 1 1 1 0 1 1 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 1 0 0 0 1 1 0 1 0 1 1 0 0 1 0
43	1 0 0 1 0 1 0 0 1 1 1 0 1 1 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 1 0 0 0 1 1 0 1 0 1 1 0 0 1
44	1 1 0 0 1 0 1 0 0 1 1 1 0 1 1 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 1 0 0 0 1 1 0 1 0 1 1 0 0

SOURCE: Adapted from Wolter (1985, p. 328) with permission of the publisher.

TABLE 4.3

Estimated Proportions Approving One Adult Hitting Another in the BRR Replicates: General Social Survey, 1984 ($n = 1{,}473$)

	Estimate (percentage)			*Estimate (percent)*	
Replicate Number	*Replicate*	*Complement*	*Replicate Number*	*Replicate*	*Complement*
1	60.9	59.2	23	61.4	58.6
2	60.1	59.9	24	57.7	62.4
3	62.1	57.9	25	60.4	59.6
4	58.5	61.7	26	61.7	58.2
5	59.0	61.0	27	59.3	60.6
6	59.8	60.2	28	62.4	57.6
7	58.5	61.5	29	61.0	58.9
8	59.0	61.0	30	61.2	58.7
9	61.3	58.8	31	60.9	59.1
10	59.2	60.8	32	61.6	58.5
11	61.7	58.3	33	61.8	58.2
12	60.2	59.8	34	60.6	59.4
13	62.1	58.7	35	58.6	61.5
14	59.7	60.4	36	59.4	60.7
15	58.1	62.0	37	59.8	60.3
16	56.0	64.2	38	62.0	58.1
17	59.8	60.3	39	58.1	61.9
18	58.6	61.3	40	59.6	60.5
19	58.9	61.1	41	58.8	61.2
20	60.8	59.3	42	59.2	60.8
21	63.4	56.5	43	58.7	61.4
22	58.3	61.7	44	60.5	59.5

Overall estimate = 60.0

Variance estimates	*Variance*	*Standard Error*	*Design Effect*
By Equation 4.3	0.000231	0.0152	1.42
By Equation 4.4	0.000227	0.0151	1.40

it was applied to computation of variance in complex surveys by Frankel (1971) in the same manner as the BRR method and named the jackknife repeated replication (JRR). As is BRR, the JRR technique generally is applied to PSUs within strata.

The basic principle of jackknifing can be illustrated by estimating sampling variance of the sample mean from a simple random sample. Suppose $n = 5$ and sample values of y are 3, 5, 2, 1, and 4. The sample mean then is $\bar{y} = 3$, and its sampling variance, ignoring the FPC, is

$$v(\bar{y}) = \frac{\sum (y_i - \bar{y})^2}{n(n-1)} = 0.5. \tag{4.5}$$

The jackknife variance of the mean is obtained as follows.

1. Compute a pseudo sample mean deleting the first sample value, which results in $\bar{y}_{(1)} = (5+2+1+4)/4 = 12/4$. Now, by deleting the second sample value instead, we obtain the second pseudo mean $\bar{y}_{(2)} = 10/4$; likewise $\bar{y}_{(3)} = 13/4$, $\bar{y}_{(4)} = 14/4$, and $\bar{y}_{(5)} = 11/4$.
2. Compute the mean of the five pseudo-values; $\bar{\bar{y}} = \sum \bar{y}_{(i)}/n = (60/4)/5 = 3$, which is the same as the sample mean.
3. The variance can then be estimated from the variability among the five pseudo-means, each of which contains four observations,

$$v(\bar{\bar{y}}) = \frac{(n-1)\sum (\bar{y}_{(i)} - \bar{\bar{y}})^2}{n} = 0.5, \tag{4.5}$$

   which gives the same result as Equation 4.5.

The replication-based procedures have a distinct advantage: They can be applied to estimators that are not expressible in terms of formulas, such as the sample median, as well as to formula-based estimators. No formula is available for the sampling variance of the median, but the jackknife procedure can offer an estimate. Using the same example as above, the sample median is 3 and the five pseudo-medians are 3, 2.5, 3.5, 3.5, and 2.5 (the mean of these pseudo-medians is 3). The variance of the median is estimated as 0.8, using Equation 4.6.

In the same manner, the jackknife procedure also can be applied to the replicated sampling. We can remove replicates one at a time and compute pseudo-values to estimate the jackknife variance, although this does not offer any computational advantage in this case. But it also can be applied to any random groups that are formed from any probability sample. For instance, a systematic sample can be divided into random or systematic subgroups for the jackknife procedure. For other sample designs, random groups can be formed following the practical rules suggested by Wolter (1985, pp. 31–33). The basic idea is to form random groups in such a way that each random group has the same sample design as the parent sample. This requires detailed information on the actual sample design, but unfortunately such information usually is not available in most public-use survey data files. The jackknife procedure is, therefore, usually applied to PSUs rather than to random groups.

For a paired selection design, the replicate is formed removing one PSU from a stratum and weighting the remaining PSU to retain the stratum's proportion in the total sample. The complement replicate is formed in the same manner by exchanging the removed and retained PSU in the stratum.

A pseudo-value is estimated from each replicate. For a weighted sample, the sample weights in the retained PSU need to be inflated to account for the observations in the removed PSU. The inflated weight is obtained by dividing the sum of the weights in the retained PSU by a factor $(1 - w_d/w_t)$, where w_d is the sum of weights in the deleted PSU and w_t is the sum of weights in all the PSUs in that stratum. The factor represents the complement of the deleted PSU's proportion of the total stratum weight. Then the variance of a sample statistic in a paired selection design calculated from the full sample can be estimated from pseudo-values u'_h and complement pseudo-values u''_h in stratum h by

$$v(\bar{u}) = \sum [(u'_h - \bar{u})^2 + (u''_h - \bar{u})^2]/2 = \sum (u'_h - u''_h)^2/4. \qquad (4.7)$$

This estimator has the same form as Equation 4.3 and can be modified to include one replicate, without averaging with the complement, from each stratum, as in Equation 4.4 for the BRR method, which gives

$$v'(\bar{u}) = \sum (u'_h - \bar{u})^2. \qquad (4.8)$$

The JRR is not restricted to a paired selection design but is applicable to any number of PSUs per stratum. If we let u_{hi} be the estimate of U from the h-th stratum and i-th replicate, n_h be the number of sampled PSUs in the h-th stratum, and r_h be the number of replicates formed in stratum h, then the variance is estimated by

$$v(\bar{u}) = \sum_h^L \left(\frac{n_h - 1}{r_h}\right) \sum_i^{r_h} (u_{hi} - \bar{u})^2. \qquad (4.9)$$

If each of the PSUs in stratum h is removed to form a replicate, $r_h = n_h$ in each stratum, but the formation of n_h replicates in h-th stratum is not required. When the number of strata is large and n_h is two or more, the computation can be reduced by using only one replicate in each stratum. However, a sufficient number of replicates must be used in analytic studies to ensure adequate degrees of freedom.

Table 4.4 shows the results of applying the JRR technique to the collapsed paired design of the 1984 GSS used in the BRR computation. Estimated proportions of adults approving "the hitting of other adults" are shown for the 42 jackknife replicates and their complements. Applying Equation 4.7, we obtain a variance estimate of 0.000238 with a design effect of 1.46, and these are about the same as those obtained by the BRR technique. Using only the 42 replicates and excluding the complements (Equation 4.8), we obtain a variance estimate of 0.000275 with a design effect of 1.68.

TABLE 4.4
Estimated Proportions Approving One Adult Hitting Another in the JRR Replicates: General Social Survey, 1984

	Estimate (percentage)			*Estimate (percent)*	
Replicate Number	*Replicate*	*Complement*	*Replicate Number*	*Replicate*	*Complement*
1	60.2	59.8	22	60.3	60.0
2	60.2	59.8	23	60.0	60.0
3	60.0	60.0	24	60.4	59.6
4	60.3	59.8	25	60.1	59.8
5	60.0	60.1	26	59.8	60.3
6	59.9	60.1	27	59.9	60.1
7	60.0	60.0	28	60.1	60.0
8	60.0	60.0	29	59.5	60.3
9	59.9	60.2	30	59.9	60.1
10	60.1	60.0	31	59.6	60.2
11	59.8	60.2	32	60.5	59.6
12	59.9	60.1	33	60.1	59.9
13	59.8	60.2	34	60.3	59.8
14	60.0	60.1	35	60.1	59.8
15	59.6	60.5	36	60.2	59.8
16	60.4	59.6	37	60.0	60.0
17	59.9	60.0	38	59.6	60.4
18	59.8	60.2	39	59.9	60.1
19	59.8	60.2	40	60.5	59.6
20	59.9	60.1	41	60.4	59.8
21	60.0	60.0	42	60.7	59.4

Overall estimate = 60.0

Variance estimates	*Variance*	*Standard Error*	*Design Effect*
By Equation 4.7	0.000238	0.0152	1.46
By Equation 4.8	0.000275	0.0166	1.68

From a closer examination of data in Table 4.4, one may get an impression that there is less variation among the JRR replicate estimates than among the BRR replicate estimates in Table 4.3. We should note, however, that the JRR represents a different strategy that uses a different method to estimate the variance. Note that Equation 4.3 for the BRR includes the number of replicates (t) in the denominator, whereas Equation 4.7 for the JRR is not dependent on the number of replicates. The reason is that in the JRR, the replicate estimates themselves are dependent on the number of replicates formed. Because the replicate is formed deleting one unit, the replicate estimate would be closer to the overall estimate when a large number of units is available to form the replicates, compared to the situation where a small number of units is used.

Therefore, there is no reason to include the number of replicates in Equations 4.7 and 4.8. However, the number of replicates needs to be taken into account when the number of replicates used is smaller than the total number of PSUs, as in Equation 4.9.

In summary, the JRR technique is based on a pseudo-replication method and can estimate sampling variances from complex sample surveys. No restrictions on the sample selection design are needed, but forming replicates requires considerable care and must take into account the sampling design of the original sample design. As noted, this detailed design information is seldom available to secondary data analysts. For instance, if more information on ultimate clusters had been available in the GSS data file, we could have formed more convincing random groups adhering more closely to actual sample design rather than applying the JRR technique to a collapsed paired design.

The Bootstrap Method

Closely related to BRR and JRR is the bootstrap method popularized by Efron (1979). The basic idea is to create replicates of the same size and structure as in the design by repeatedly resampling the PSUs in the observed data. Applying the bootstrap method to 84 PSUs in 42 pseudo-strata in the GSS data, one would sample 84 PSUs (using a *with replacement* sampling procedure), two from each stratum. In some strata, the same PSU may be selected twice. The sampling is repeated a large number of times, a minimum of 200 (referred to as B) times (Efron & Tibshirani, 1993, sec. 6.4). However, a much larger number of replications usually is required to get a less variable estimate (Korn & Graubard, 1999, p. 33). For each replicate created (u'_i), the parameter estimate is calculated. Then the bootstrap estimate of the variance of the mean of all replicate estimates is given by

$$v(\bar{u}) = \frac{1}{B}\sum_{i=1}^{B}(u'_i - \bar{u})^2. \tag{4.10}$$

This estimator needs to be corrected for bias by multiplying it by $(n-1)/n$. When n is small, the bias can be substantial. In our example, there are two PSUS in each stratum, and the estimated variance needs to be halved. An alternative approach to correct the bias is to resample $(n_h - 1)$ PSUs in stratum h and multiply the sample weights of the observations in the resampled PSUs by $n_h/(n_h - 1)$ (Efron, 1982, pp. 62–63). In our example, this would produce half-sample replicates as in BRR. The bootstrap estimate based on at least 200 replicates would then be about the same as the BBR estimate based on 44 half-sample replicates. Because of the large number of replications required in the bootstrap method, this method has not yet been used extensively for variance estimation in complex survey analysis.

Various procedures of applying the bootstrap method for variance estimation and other purposes have been suggested (Kovar, Rao, & Wu, 1988; Rao & Wu, 1988; Sitter, 1992). Although the basic methodology is widely known, many different competing procedures have emerged in selecting bootstrap samples. For example, Chao and Lo (1985) suggested duplicating each observation in the host sample N/n times to create the bootstrap population for simple random sampling without replacement. For sampling plans with unequal probability of selection, the replication of the observations needs to be proportionate to the sample weight; that is, the bootstrap sample should be selected using the PPS procedure. These options and the possible effects of deviating from the fundamental assumption of independent and identically distributed samples have not been thoroughly investigated.

Although it is promising for handling many statistical problems, the bootstrap method appears less practical than BRR and JRR for estimating the variance in complex surveys, because it requires such a large number of replicates. Although BRR and JRR would produce the same results when applied by different users, the bootstrap results may vary for different users and at different tries by the same user, because the replication procedure is likely to yield different results each time. As Tukey (1986, p. 72) put it, ‘‘For the moment, jackknifing seems the most nearly realistic approach to assessing many of the sources of uncertainty’’ when compared with bootstrapping and other simulation methods. The bootstrap method is not implemented in the available software packages for complex survey analysis at this time, although it is widely used in other areas of statistical computing.

The Taylor Series Method (Linearization)

The Taylor series expansion has been used in a variety of situations in mathematics and statistics. One early application of the series expansion was to obtain an approximation to the value of functions that are hard to calculate, for example, the exponential e^x or logarithmic [log(x)] function. This application was in the days before calculators had special function keys and when we did not have access to the appropriate tables. The Taylor series expansion for e^x involves taking the first- and higher-order derivatives of e^x with respect to x; evaluating the derivatives at some value, usually zero; and building up a series of terms based on the derivatives. The expansion for e^x is

$$1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots$$

This is a specific application of the following general formula expanded at a:

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)(x-a)^2}{2!} + \frac{f'''(a)(x-a)^3}{3!} + \ldots$$

In statistics, the Taylor series is used to obtain an approximation to some nonlinear function, and then the variance of the function is based on the Taylor series approximation to the function. Often, the approximation provides a reasonable estimate to the function, and sometimes the approximation is even a linear function. This idea of variance estimation has several names in the literature, including the linearization method, the delta method (Kalton, 1983, p. 44), and the propagation of variance (Kish, 1965, p. 583).

In statistical applications, the expansion is evaluated at the mean or expected value of x, written as $E(x)$. If we use $E(x)$ for a in the above general expansion formula, we have

$$f(x) = f[E(x)] + f'[E(x)][x - E(x)] + f''[E(x)][x - E(x)]^2/2! + \ldots$$

The variance of $f(x)$ is $V[f(x)] = E[f^2(x)] - E^2[f(x)]$ by definition, and using the Taylor series expansion, we have

$$V[f(x)] = \{f'[E(x)]\}^2 V(x) + \ldots \tag{4.11}$$

The same ideas carry over to functions of more than one random variable. In the case of a function of two variances, the Taylor series expansion yields

$$V[f(x_1, x_2)] \cong \left(\frac{\partial f}{\partial x_1}\right)\left(\frac{\partial f}{\partial x_2}\right) Cov(x_1, x_2) \tag{4.12}$$

Applying Equation 4.12 to a ratio of two variables x and y—that is, $r = y/x$—we obtain the variance formula for a ratio estimator

$$\begin{aligned} V(r) &= \frac{V(y) + r^2 V(x) - 2r\ Cov(x, y)}{x^2} + \ldots \\ &= r^2\left(\frac{V(y)}{y^2} + \frac{V(x)}{x^2} - \frac{2\ Cov(x, y)}{xy}\right) + \ldots \end{aligned}$$

Extending Equation 4.12 to the case of c random variables, the approximate variance of $\theta = f(x_1, x_2, \ldots, x_c)$ is

$$V(\theta) \cong \sum\sum \left(\frac{\partial f}{\partial x_i}\right)\left(\frac{\partial f}{\partial x_j}\right) Cov(x_i, x_j) \tag{4.13}$$

TABLE 4.5
Standard Errors Estimated by Taylor Series Method for Percentage Approving One Adult Hitting Another: General Social Survey, 1984 ($n = 1{,}473$)

	Subgroup	*Estimate (percent)*	*Standard Error (percent)*	*Design Effect*
Overall		60.0	1.52	1.41
Gender	Male	63.5	2.29	1.58
	Female	56.8	1.96	1.21
Race	White	63.3	1.61	1.43
	Non white	39.1	3.93	1.30
Education	Some college	68.7	2.80	1.06
	High school graduate	63.3	2.14	1.55
	All others	46.8	2.85	1.27

Applying Equation 4.13 to a weighted estimator

$$f(Y) = \hat{Y}_i = \sum w_i y_{ij},\ j = 1, 2, \ldots, c$$

involving c variables in a sample of n observations, Woodruff (1971) showed that

$$V(\theta) \cong V\left[\sum w_i \sum\left(\frac{\partial f}{\partial y_j}\right) y_{ij}\right]. \qquad (4.14)$$

This alternative form of the linearized variance of a nonlinear estimator offers computational advantages because it bypasses the computation of the $c \times c$ covariance matrix in Equation 4.13. This convenience of converting a multistage estimation problem into a univariate problem is realized by a simple interchange of summations. This general computational procedure can be applied to a variety of nonlinear estimators, including regression coefficients (Fuller, 1975; Tepping, 1968).

For a complex survey, this method of approximation is applied to PSU totals within the stratum. That is, the variance estimate is a weighted combination of the variation in Equation 4.14 across PSUs within the same stratum. These formulas are complex but can require much less computing time than the replication methods discussed above. This method can be applied to any statistic that is expressed mathematically—for example, the mean or the regression coefficient—but not to such nonfunctional statistics as the median and other percentiles.

We now return to the GSS example of estimating the variance of sample proportions. Table 4.5 shows the results of applying the Taylor series method to the proportion of adults approving the hitting of other adults, analyzed by gender, race, and level of education. The proportion is computed as a ratio of weighted sums of all positive responses to the sum of all the weights. Its standard error is computed applying Equation 4.14 modified to include the PSUs and strata. The design effect for the overall proportion is 1.41, which is about the same as those estimated by using the other two methods, whose results are shown in Tables 4.3 (BBR) and 4.4 (JRR). The estimated proportion varies by gender, race, and the level of education. Because the subgroup sizes are small, the standard errors for the subgroups are larger than that for the overall estimate. In addition, the design effects for subgroup proportions are different from that for the overall estimate.

In this chapter, we presented several methods of estimating variance for statistics from complex surveys (for further discussion, see Rust and Rao, [1996]). Examples from GSS and other surveys tend to show that the design effect is greater than 1 in most complex surveys. Additional examples can be found in E. S. Lee, Forthofer, and Lorimor (1986) and Eltinge, Parsons, and Jang (1997). Examples in Chapter 6 will demonstrate the importance of using one of the methods reviewed above in the analysis of complex survey data.

5. PREPARING FOR SURVEY DATA ANALYSIS

The preceding chapters have concentrated on the complexity of survey designs and techniques for variance estimation for these designs. Before applying the sample weights and the methods for assessing the design effect, one must understand the survey design and the data requirements for the estimation of the statistics and the software intended to be used. These requirements are somewhat more stringent for complex survey data than for data from an SRS because of the weights and design features used in surveys.

Data Requirements for Survey Analysis

As discussed in Chapter 3, the weight and the design effect are basic ingredients needed for a proper analysis of survey data. In preparing for an analysis of survey data from a secondary source, it is necessary to include the weights and the identification of sampling units and strata in the working data file, in addition to the variables of interest. Because these design-related data items are labeled differently in various survey data sources, it is important to read the documentation or consult with the source agency or person to understand the survey design and the data preparation procedures.

The weights usually are available in major survey data sources. As noted earlier, the weights reflect the selection probabilities and the adjustments