EXAMPLE 11.2 An economist wants to estimate the average weekly amount spent on food by families with children in a certain county known to be a poverty area. A complete list of all the 250 families in the county is available, but identifying those families with children is impossible. The economist selects a simple random sample of $n = 50$ families and finds that $n_1 = 42$ families have at least one child. The 42 families with children are interviewed and give the following information:

$$\sum_{j=1}^{42} y_{1j} = \$1720 \qquad s_1^2 = 42.975$$

Estimate the average weekly amount spent on food by all families with children and place a bound on the error of estimation.

SOLUTION The estimator of the population mean is $\bar{y}_1$, given by Eq. (11.5). Calculations yield

$$\bar{y}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} y_{1j} = \frac{1}{42}(1720) = 40.95$$

The quantity $1 - n_1/N_1$ must be estimated by $1 - n/N$ because N_1 is unknown. The estimated variance of $\bar{y}_1$, given in Eq. (11.6), then becomes

$$\hat{V}(\bar{y}_1) = \left(1 - \frac{n}{N}\right)\frac{s_1^2}{n_1} = \left(1 - \frac{50}{250}\right)\left(\frac{42.975}{42}\right)$$
$$= 0.818$$

Thus, the estimate of the population average, with a bound on the error of estimation, is given by

$$\bar{y}_1 \pm 2\sqrt{\hat{V}(\bar{y}_1)} \quad \text{or} \quad 40.95 \pm 2\sqrt{0.818}$$
$$\text{or} \quad 40.95 \pm 1.81$$

Our best estimate of the average weekly amount spent on food by families with children is \$40.95. The error of estimation should be less than \$1.81 with a probability of approximately .95. ■

EXAMPLE 11.3 A preliminary study of the county in Example 11.2 reveals $N_1 = 205$ families with children. Using this information and data given in Example 11.2, estimate the total weekly amount spent on food by families with children. (Note: N_1 will vary over time. We assume that the value of N_1 used in this analysis is correct.)

SOLUTION The best estimator of the total is $N_1\bar{y}_1$ given in Eq. (11.7), which yields an estimate of

$$N_1\bar{y}_1 = 205(40.95) = 8394.75$$

The estimated variance of $N_1\bar{y}_1$ is then, from Eq. (11.8),

$$\hat{V}(N_1\bar{y}_1) = N_1^2\left(1 - \frac{n_1}{N_1}\right)\frac{s_1^2}{n_1}$$
$$= (205)^2\left(1 - \frac{42}{205}\right)\left[\frac{42.975}{42}\right] = 34{,}191.19$$

The estimate of the total weekly amount that families with children spend on food, given with a bound on the error of estimation, is

$$N_1\bar{y}_1 \pm 2\sqrt{\hat{V}(N_1\bar{y}_1)} \quad \text{or} \quad 8394.75 \pm 2\sqrt{34{,}191.19}$$
$$\text{or} \quad 8394.75 \pm 369.82 \qquad ■$$

EXAMPLE 11.4 Suppose that the experimenter in Example 11.3 doubts the accuracy of the preliminary value of N_1. Use the data from Example 11.3 to estimate the total weekly amount spent on food by families with children, without using the value given for N_1.

SOLUTION The estimator of the total that does not depend on N_1 is $\hat{\tau}_1$, given by Eq. (11.9). Thus,

$$\hat{\tau}_1 = \frac{N}{n}\sum_{j=1}^{n_1} y_{1j} = \frac{250}{50}(1720) = 8600$$

The adjustment to the sample variance that comes about by adding 8 zeros to the 42 measurements used above is a bit tricky here because we do not have the data. You may recall that, in general,

$$s^2 = \frac{1}{n-1}\left[\sum y_i^2 - n(\bar{y})^2\right]$$

The sample mean of the new sample of 50 can be calculated easily because the sample total of 1720 does not change. The sum of squares of the measurements does not change either because adding zeros contributes zero to the result. From the knowledge of the sample variance for the 42 data values, it can be ascertained that the sum of squares term is approximately 72,200. Then, $s_n^2 = 265.96$. Substituting into Eq. (11.10) gives the estimated variance of $\hat{\tau}_1$:

$$\begin{aligned}\hat{V}(\hat{\tau}_1) &= N^2\left(1 - \frac{n}{N}\right)\frac{s_n^2}{n} \\ &= (250)^2\left(1 - \frac{50}{250}\right)\left(\frac{265.96}{50}\right) \\ &= 265{,}960\end{aligned}$$

Thus, the estimate of the total weekly amount spent on food, with a bound on the error of estimation, is

$$\hat{\tau}_1 \pm 2\sqrt{\hat{V}(\hat{\tau}_1)} \quad \text{or} \quad 8600 \pm 2\sqrt{265{,}960}$$
$$\text{or} \quad 8600 \pm 1031.44$$

The estimate has a large bound on the error of estimation, which could be reduced by increasing the sample size n. ■

Из књиге “Elementary Survey Sampling”, R. L. Scheaffer, W. Mendenhall III, R. L. Ott, K. Gerow

```
#популација
farme <- c("A", "B", "C")
#вредности обележја од интереса на свим јединицама у популацији
proizvodnja_psenice <- c(11, 6, 25) #у тонама

(populacijski_total <- sum(proizvodnja_psenice)) #непознато!
[1] 42

#бира се случајан узорак са понављањем обима 2 и вероватноћама одабира
> verovatnoce_odabira <- c(0.3, 0.2, 0.5)
#вероватноће укључења првог реда су:
[1] 0.51 0.36 0.75

#сви могући узорци, као неуређени скупови кардиналности 2 са евентуалним понављањима
ознака јединица (има их 3+3) и вероватноће избора сваког од њих су:
     [,1] [,2] [,3]
 [1,] "A"  "A"  0.09
 [2,] "A"  "B"  0.12
 [3,] "A"  "C"  0.30
 [4,] "B"  "B"  0.04
 [5,] "B"  "C"  0.20
 [6,] "C"  "C"  0.25

#ажурирани подаци са оценама (Hansen-Hurwitz, Horvitz-Thompson, Hajèk, редом) тотала за
сваки од узорака
podaci
     [,1] [,2] [,3] [,4] [,5] [,6]
 [1,] "A"  "A"  0.09  36.67  21.57  33.00
 [2,] "A"  "B"  0.12  33.33  38.24  24.21
 [3,] "A"  "C"  0.30  43.33  54.90  50.00
 [4,] "B"  "B"  0.04  30.00  16.67  18.00
 [5,] "B"  "C"  0.20  40.00  50.00  36.49
 [6,] "C"  "C"  0.25  50.00  33.33  75.00

#провера својства непристрасности предложених оцена у односу на метод одабира узорка
mat_ocekivanje
[1] 41.9989  41.9994  47.6432

#средње квадратне грешке предложених оцена у односу на метод одабира узорка
[1] 34.66774  146.4407  365.8301 #непознато!
```