Example 4.9. Example 3.2 displayed estimates for a stratified random sample from the Census of Agriculture population. The stratified sample, taken with proportional allocation, produced estimates with smaller variances than the SRS in Example 2.6.

But what if you took an SRS and only later realized that you should have taken a stratified sample? Or if you did not have region membership available for the counties in the sampling frame? Let's poststratify the SRS from Example 2.6 and find out. The quantities needed for the calculation are given in Table 4.4.

TABLE 4.4
Weight adjustments for poststratification in Example 4.9. The last two columns, $\bar{y}_h$ and s_h, give the poststratum mean and standard deviation, respectively.

Region	N_h	n_h	$\hat{N}_h$	w_i	w_i^*	$\bar{y}_h$	s_h
Northeast	220	24	246.24	10.26	9.1667	71970.83	65000.06
North Central	1054	107	1097.82	10.26	9.8505	350292.01	294715.13
South	1382	130	1333.80	10.26	10.6308	206246.35	277433.61
West	422	39	400.14	10.26	10.8205	598680.59	516157.67
Total	3078	300	3078.00				

The poststratification-adjusted weights w_i^* differ from the original sampling weights $w_i = 3078/300 = 10.26$. The poststratified weight for every county in the Northeast, where the SRS contained more units than would have been drawn in a stratified sample with proportional allocation, is

$$w_i^* = \frac{N_{\text{Northeast}}}{\hat{N}_{\text{Northeast}}} w_i = \frac{220}{246.24}(10.26) = 9.1667.$$

The poststratified weight for Northeast counties is smaller than 10.26 to account for the fact that the randomly selected sample contained, by chance, more counties in the poststratum than its share of the population. The poststratified weights in the West poststratum are larger than 10.26 to correct for the sample having more Western counties than it would under proportional allocation.

The weight adjustments force the estimated counts from each poststratum to equal the true poststratum count, N_h. Thus, $\sum_{i \in \mathcal{S}} w_i^* x_{i1} = (24)(9.1667) = 220$, $\sum_{i \in \mathcal{S}} w_i^* x_{i2} = 1054$, $\sum_{i \in \mathcal{S}} w_i^* x_{i3} = 1382$, and $\sum_{i \in \mathcal{S}} w_i^* x_{i4} = 422$.

The poststratified estimate of the population mean is

$$\bar{y}_{\text{post}} = \frac{\sum_{i \in \mathcal{S}} w_i^* y_i}{\sum_{i \in \mathcal{S}} w_i^*} = \frac{15833583 + 369207778.5 + 285032455.7 + 252643209}{3078} = 299{,}778.$$

From (4.27), the standard error of $\bar{y}_{\text{post}}$ is

$$\text{SE}\left(\bar{y}_{\text{post}}\right) = \sqrt{\left(1 - \frac{300}{3078}\right) \sum_{h=1}^{H} \frac{N_h}{3078} \frac{s_h^2}{300}} = 17{,}443.$$

By contrast, the standard error of the sample mean, $\bar{y}$, from Example 2.7, is 18,898. The poststratification reduces the standard error because the weighted average of the within-poststratum variances, $\sum_{h=1}^{H}(N_h/N)s_h^2$, is smaller than s^2. ■

Difference between stratification and poststratification. In both stratification and post-stratification, each population member belongs to exactly one of H possible groups. In stratified random sampling, independent SRSs are selected from each of the H groups