Solutions to Exercises

Chapter 1

1.1 If a public gathering can be considered to be a representative sample of the entire population, estimates derived from the show of hands or applause can provide valid estimates.

1.3 (1) The people contacted should be representative of the population.
(2) Persons should be contacted during the peak periods and also other times during the week.

1.5 Proportional distribution of the sample to the four classes provides better estimates; see Chapter 5.

1.7 Systematic sampling should be recommended for the characteristics with an approximate linear trend. Otherwise, both systematic and simple random sampling provide almost the same precision.

1.9 (1) The purchases for these items should be large in number.
(2) The returned questionnaires should be representative of all the consumers who purchased the products.

1.11 (1) To estimate the percentage of the purchases, add the numbers of purchases in the two groups, subtract the number of purchases from the 30 duplicated units from the total, and divide the resulting number by 270.
(2) Similarly, to estimate the average of the purchases, add the totals for the two groups, subtract the amount for the 30 duplicated units, and divide the resulting amount by 270.

Chapter 2

2.1 (a) For the 80 corporations, $\hat{Y} = 20{,}000$.
(b) S.E. $(\hat{Y} = 1{,}704)$.
(c) 95% confidence limits are (16660, 23340).

2.3 (a) S.E. of the mean is $15/(30)^{1/2} = 2.74$, $t_{29}(.05) = 2.0452$, and 95% confidence limits are (29.4, 40.6).
(b) C.V. of the sample mean is equal to 2.74/35 = 0.09, that is, 9%.

2.5 (a) $V(t_1) = [Nn_2/(N_1 + n_2)]^2 V(\bar{y}_2)$. From the sample, $v(\bar{y}_2) = 20$, $v(t_1) = 20{,}663.7$ and S.E.$(t_1) = 143.75$.
(b) $E(t_1) = N(Y_1 + n_2\bar{Y}_2)/(N_1 + n_2)$, which is not equal to $Y = Y_1 + Y_2$, and hence t_1 is biased for Y.

2.7 (a) $n = 32$
(b) $n = 43$
(c) $n = 318$.

2.9 For the three types of households, the required sample sizes are (1) $n = 136$, (2) $n = 169$, and (3) $n = 211$.

2.11 The result in (c) follows from noting that a pair of units appear in ${}_{N-2}C_{n-2}$ of the ${}_{N}C_{n}$ samples.

Chapter 3

3.1 For the sample sizes 150 and 300, the variances are $V_1 = (3000 - 150)S^2/(3000 \times 50) = 63.33S^2/10^4$ and $V_2 = (3000 - 300)S^2/(3000 \times 300) = 30S^2/10^4$.
(a) Relative decrease in the variance is $(V_1 - V_2)/V_1$, which is 53%.
(b) The S.E.s respectively are equal to $7.96S/100$ and $5.48S/100$, and hence the relative decrease in the S.E. is equal to $(7.96 - 5.48)/7.96 = 0.31$, that is, 31%.
(c) 31%, as in (b).
(d) Relative increase in the precision is $(V_1/V_2) - 1 = 1.11$, that is, 111%.
(e) Confidence widths for the two cases are $2Z(7.96S)/100$ and $2Z(5.48S)/100$. Relative decrease in the confidence width is equal to 0.31, that is, 31%.
(f) Same as (a)–(e) for the estimation of the total.

3.3 (a) $d = \bar{Y}_1 - \bar{y}_2 = 50$.
(b) $V(d) = V(\bar{y}_2) = 453.75$ and hence S.E.$(d) = 21.3$.

(c) The 95% confidence limits for $\bar{Y}_1 - \bar{Y}_2$ are (8.25, 91.75).
(d) Since $E(\bar{Y}_1 - \bar{y}_2) = \bar{Y}_1 - \bar{Y}_2$, it is unbiased.

3.5 (a) 20.
(b) 30.91.
(c) -40.58, 80.58

3.7 First list: simple random sampling can be recommended for all the three items.
Second list: systematic sampling for (1) and (2), and simple random sampling for the third list can be recommended.
Third list: same as for the second list.

3.9 (1) Bias, variance, and MSE of $\bar{y}_1$ are given by $B(\bar{y}_1) = \bar{Y}_1 - \bar{Y} = W_2(\bar{Y}_1 - \bar{Y}_2), V(\bar{y}_1) = (N_1 - n_1)S_1^2/N_1 n_1$ and $\text{MSE}(\bar{y}_1) = V(\bar{y}_1) + B^2(\bar{y}_1)$. The bias will be small if $\bar{Y}_2$ is close to $\bar{Y}_1$, and the S.E. will be small if n_1 is large.
(2) Same as (1).
(3) Since the expected value of the resulting estimator is $W_1\bar{Y}_1$, it has a bias of $W_1\bar{Y}_1 - \bar{Y} = -W_2\bar{Y}_2$, which will be small if the size of the nonrespondents is not large. This estimator has a variance of $(n_1/n)^2(N_1 - n_1)S_1^2 / N_1 n_1$, which will be small if n is large.

3.11 (a) For the mean of the systolic pressure, 90% confidence limits are (152.8, 166.7).
(b) The hypothesis that the mean is greater than 165 is not rejected at the 5% level of significance.

3.13 The population C.V. of y is $S_y/\bar{Y} = 60.66/670 = 0.09$, that is 9%. For the sample C.V., $s_y/\bar{y}$, the expectation, bias, variance, and MSE, respectively, are equal to 0.08, 0.01, 0.0018, and 0.0019.

The population correlation of x and y is $\rho = 0.41$. For the sample correlation coefficient, the expectation, bias, variance, and MSE are equal to 0.39, 0.02, 0.3853, and 0.3857.

3.15 (a) From the Cauchy-Schwartz inequality, $S_{xy}^2 \le S_x^2 S_y^2$, and hence $\rho^2 \le 1$, that is, $-1 \le \rho \le 1$.
(b) Similarly, $s_{xy}^2 \le s_x^2 s_y^2$, and hence $r^2 \le 1$, that is, $-1 \le r \le 1$.

3.17 (c) The unbiasedness follows from (b) by noting that the probability of selecting a pair of units is $n(n-1)/N(N-1)$.

3.19 (a) Since each of the population units is selected with equal probability $p_i = 1/N$, $E(y_i) = \sum_1^N p_i y_i = \sum_1^N y_i/N = \bar{Y}$.
(b) $V(y_i) = E(y_i - \bar{Y})^2 = \sum_1^N p_i(y_i - \bar{Y})^2 = (N-1)S^2/N = \sigma^2$.

(c) Since the units are replaced, they are selected independently. Hence, $\mathrm{Cov}(y_i, y_j) = E(y_i - \overline{Y})(y_j - \overline{Y}) = E(y_i - \overline{Y}) \times E(y_j - \overline{Y}) = 0$. This result can also be obtained by noting that $E(y_i y_j) = E(y_i)E(y_j) = \overline{Y}^2$.

Chapter 4

4.1 (a) Public institutions: $p = 6/10$, S.E.$(p) = 0.1456$, $\hat{C} = 30$, and S.E.$(\hat{C}) = 7$.
Private institutions: $p = 1/10$, S.E.$(p) = 0.09$, $\hat{C} = 5$, and S.E.$(\hat{C}) = 4$.
(b) Estimate of the difference in the proportions for the two types of institutions is $d = 0.5$ and S.E. $(d) = 0.145$.
Estimate for the difference in the numbers of institutions with enrollments exceeding 100,000 is $49(.5) = 24.5$, that is, 25, and this estimate has a S.E. of $49(.145) = 7$.

4.3 For the first poll, $e^2 = (1.96)^2(.49)(.51)/597 = 0.0016$. Hence $e = 0.04$, that is, 4%.
For the second poll, $e^2 = (1.96)^2(.44)(.56)/597 = 0.0016$, and hence $e = 0.04$ again.
For the last two polls, $e^2 = (1.96)^2(.57)(.43)/500 = 0.0019$. Hence $e = 0.044$, that is, close to 4.5%.

4.5 (a) $p = 0.31 + 0.12 = .43$, that is 43%.
(b) $v(p) = (.43)(.57)/499 = 0.0005$. Hence S.E$(p) = 0.022$, that is 2.2%.
(c) The 95% confidence limits are (39, 47)%.
(d) Denoting the proportions in favor and not in favor by P and $Q = 1 - P$, their difference is $D = 2P - 1$. The null and alternative hypotheses are $D \leq .10$ and $D > .10$.
From the sample results, an unbiased estimate of this difference is $2p - 1 = .14$, which has the sample variance $v(2p - 1) = 4v(p) = 4(.57)(.43)/499 = 0.002$ and hence the S.E. of 0.045.
Now, $Z = (.14 - .10)/.045 = 0.89$, which is less than 1.65, the percentile of the standard normal distribution for the one-sided 5% level of significance. Hence, we do not reject the null hypothesis that $(P - Q)$ is less than 10%.

4.7 (a) For the Democrats, $p = .21 + .22 + .31 = .74$, that is, 74%. Since $v(p) = (.74)(.26)/1697 = .00011$, S.E. $(p) = .01$, that is, 1%. The 95% confidence limits are (72, 76)%.

(b) For the Republicans, $p = 70\%$, S.E.$(p) = 1.2\%$, and 95% confidence limits are (67.6, 72.4).

(c) For the Democrats and Republicans together, $p = [1698(.74) + 1432(.70)/3130 = .722$, that is, 72.2%. Now, $v(p) = (.722)(.278)/3129 = 64 \times 10^{-6}$, and hence S.E.$(p) = 0.8\%$. 95% confidence limits for this percentage are (70.6, 73.8)%.

d) An estimate for the percentage without college education is $100 - 72.2 = 27.8\%$. It has the same S.E. of 0.8 as in (c). Confidence limits for this percentage are (26.2, 29.4).

4.9 (a) For the Freshman–Sophomore, Junior, and Senior groups, the error e is $.2/5 = .04$, $.5/5 = .1$, and $.8/5 = .16$, respectively.

The required sample sizes are 322, 88, and 24, respectively. For the first group, the small value of .04 for the error resulted in a relatively large sample size.

(b) With $e = .10$, same for each group, the required sample sizes for the three groups are 60, 88, and 59. Since $p = .5$ for the Juniors, sample size required for this group is relatively larger.

4.11 (a) 82.

(b) 197.

(c) 188.

4.13 Preliminary estimates of the proportions in favor of continuing the program are $p_1 = 8/11$, $p_2 = 12/25$, and $p_3 = 1/2$.

(a) $V(p_1 - p_2) = V(p_1) + V(p_2)$. Replacing the observed sample sizes n_1 and n_2 in this expression by their expected values nW_1 and nW_2, where $W_1 = N_1/N$ and $W_2 = N_2/N$, this variance approximately becomes equal to $[(N - n)/Nn][P_1Q_1/W_1 + P_2Q_2/W_2]$. Since this variance should not exceed $(.07)^2$, $n = 698$.

(b) As in (a), sample sizes required to estimate $(P_1 - P_3)$ and $(P_2 - P_3)$ with the S.E.s of the estimates not exceeding 0.07, the required sample sizes are 1001 and 280.

A sample of size 1001 from the 8500 units is required to estimate the three differences of the percentages with the S.E. not exceeding 7% in each case.

4.15 (a) From (4.20), $E(c_1c_2) = n(n - 1)C_1C_2/N(N - 1)$. Now, $\text{Cov}(c_1, c_2) = E(c_1c_2) - n^2P_1P_2 = -[(N - n)n/(N - 1)]P_1P_2$. Hence, $\text{Cov}(p_1, p_2) = -[(N - n)/(N - 1)n]P_1P_2$.

(b) An unbiased estimator for P_1P_2 is given by $(N - 1) \times np_1p_2/N(n - 1)$. Substituting this expression in the above covariance, its unbiased estimator is given by $-[(N - n)/N(n - 1)]p_1p_2$.

Chapter 5

5.1 Difference of the means: (a) 10.32.
(b) 3.81.
(c) (2.92, 17.86).
Difference of the totals: (a) 99.03.
(b) 42.88.
(c) (14.99, 183.1).

5.3 For a simple random sample of size 16 from the 43 units, $V(\bar{y}) = 1.243$.

For proportional allocation, $n_1 = 12$ and $n_2 = 4$, $V_P = 0.393$, and the relative precision is $1.243/.393 = 3.16$.

For Neyman allocation, $n_1 = 5$ and $n_2 = 11$. We select five units from the first stratum and *all* the 11 units of the second stratum. This is an example of 100% sampling from a stratum.

From (5.11), the estimator for the population mean now can be expressed as $\hat{\bar{Y}}_{\text{st}} = W_1\bar{y}_1 + W_2\bar{Y}_2$. Thus, $V_N = V(\hat{\bar{Y}}_{\text{st}}) = W_1^2V(\bar{y}_1) = 0.074$, and the precision of Neyman allocation relative to simple random sampling is $1.243/.074 = 16.80$.

5.5 In Table T2, math scores are below 600 for the eight units (4, 5, 9, 23, 24, 25, 26, 27) and 600 or more for the remaining 22 units. For the total score, the variances of the first and second stratum, respectively, are 11541.07 and 8872.73.

When $n = 10$, proportional allocation results in $n_1 = 3$ and $n_2 = 7$. With these sample sizes, from (5.12), $V(\hat{\bar{Y}}_{\text{st}}) = 635.74$. Neyman allocation results in the same sample sizes for the strata, and hence $V(\hat{\bar{Y}}_{\text{st}}) = 635.74$.

From (5.23) and (5.26), $V_P = 638.95$, and $V_N = 635.60$. The sample sizes are rounded off for finding the actual variance $V(\hat{\bar{Y}}_{\text{st}})$ from (5.12). Since the expressions for V_P and V_N ignore the rounding off, their values are slightly different from 635.74.

5.7 Since $W_1 = 2/10$, $W_2 = 8/10$, $W_1S_1^2 + W_2S_2^2 = 3060$. As described in Section 5.13, $n = 30$. Proportional allocation results in $n_1 = 6$ and $n_2 = 24$.

5.9 Proportional allocation: $n = 30$. $n_1 = 5$, $n_2 = 7$, and $n_3 = 18$.
Neyman allocation: $n = 29$. $n_1 = 3$, $n_2 = 5$, and $n_3 = 21$.

5.11 $n = 97$. $n_1 = 9$, $n_2 = 18$, and $n_3 = 70$.
$V_{\text{opt}} = 0.01424$ and S.E.$(\hat{\bar{Y}}_{\text{st}}) = 0.12$.

5.15 (a) Denote the probability of observing the n_g units by P_g. With $a_g = (P_g/n_g)^{1/2}$ and $b_g = (P_g n_g)^{1/2}$, from the Cauchy-Schwartz inequality, $(\Sigma P_g/n_g)(\Sigma P_g n_g) \geq (\Sigma p_g)^2 = 1$. This final result is the same as $E(1/n_g) \geq 1/E(n_g)$.
(b) Since the covariance of n_g and $1/n_g$ is negative, $E(n_g) \times (1/n_g) - E(n_g)E(1/n_g) \leq 0$. Hence $1 - E(n_g)E(1/n_g) \leq 0$, that is, $E(1/n_g) \geq 1/E(n_g)$.
(c) This result follows since $1/n_g$ is replaced by $1/E(n_g)$ for the average variance.

Chapter 6

6.1 (a) $N_2 = 67$, $\hat{Y}_2 = 25199$, and S.E.$(\hat{Y}_2) = 3914$.
(b) $\hat{N}_2 = 60$, $\hat{Y}_2 = 22566$, and S.E.$(\hat{Y}_2) = 3464$.

6.3 (a) and (b)

Estimates and S.E.s

	Sizes	Proportion	Total
Owners	Known		
	8549	.20, .0254	1710, 217
	Estimated		
	8897	.20, .0254	1779, 226
Renters	Known		
	6157	.158, .0287	973, 177
	Estimated		
	5809	.158, .0287	918, 167
Diff.	Known	.04, .0385	737, 280
	Estimated	.04, .0384	861, 281

(c) d = (58 - 42)/158 = 0.0253, and S.E.(d) = 0.0124.

6.5 $\bar{y} = \Sigma f_i y_i / n = 962/400 = 2.4$
$s^2 = \Sigma f_i (y_i - \bar{y})^2/(n-1) = 1.5$
$v(\bar{y}) = [(14706 - 400)/14706 \times 400](1.5) = 0.0036$, and hence S.E.$(\bar{y}) = 0.06$.

6.7 For the males, the sample mean is $\bar{y}_M = [5(3.4) + 3(4.67) + 4(4)] = 3.92$. With the expression similar to (5.6), the sample variance for the males is $s_M^2 = 1.69$. Now, $v(\bar{y}_M) = 0.14$ and hence S.E.$(\bar{y}_M) = 0.3742$.

Similarly, for the females, $\bar{y}_F = 2$, $s_F^2 = 0.64$, $v(\bar{y}_F) = 0.08$, and S.E.$(\bar{y}_F) = 0.28$.

Now, $\bar{y}_M - \bar{y}_F = 1.92$, $v(\bar{y}_M - \bar{y}_F) = 0.14 + 0.08 = 0.22$, and hence S.E.$(\bar{y}_M - \bar{y}_F) = 0.47$.

6.9 (a) For the women, the estimate is 43% with a S.E. of 1.6%. For the men, the estimate is 9% with a S.E. of 1.0%.
(b) Estimate for the difference of the percentages of the women and men is 34% with a S.E. of 1.8%.

6.11 (a) The responses now refer to the 850 women and 850 men. In this case, variance of the percentage for the women is (9.43)(.57)/849 = 288×10^{-6}, and hence it has a S.E. of 1.7%.

Variance of the percentage for the men is 98×10^{-6} and hence its S.E. is close to 1.0%.

(b) The variance of the difference of the percentages is 384×10^{-6} and hence it has a S.E. of 1.96%.

The estimates and the S.E.s in this case do not differ much from those in Exercise 6.9, since the sample size is large and a large number, 85%, of the sampled units responded.

Chapter 7

7.1 For the establishments, $\bar{\bar{Y}} = 3.02$, $S_b^2 = 6(21.0493)/8 = 15.79$.

From (7.11), $V(\bar{\bar{y}}) = [(9 - 3)/27](15.79/6) = 0.5848$ and hence S.E.$(\bar{\bar{y}}) = 0.76$.

For a simple random sample of 18 counties, $V(\bar{\bar{y}}) = [(54 - 18)/54 \times 8](15.94) = 0.5904$ and hence S.E.$(\bar{\bar{y}}) = 0.77$.

Both cluster sampling and simple random sampling have almost the same precision.

7.3 *Employment*:

For the sample clusters (1, 7, 9), from Table 7.1, $\bar{\bar{y}} = (74.2 + 38.5 + 76.3)/3 = 63$ and $s_b^2 = 2707.4$. From (7.15), $v(\bar{\bar{y}}) = (6/27)(2707.7416 = 100.29$.

From Table 7.2, $s_W^2 = (17587.70 + 2378.86 + 2876.83)/3 = 7614.46$. Now, from Appendix A7.2, $s^2 = (15/17)(7614.46) + (2/17)(2707.74) = 7037.20$. For a simple random sample of 18 clusters, $v(\bar{\bar{y}}) = [(54 - 18)/54 \times 18](7037.20) = 260.64$.

Precision of cluster sampling relative to simple random sampling is 260.64/100.29 = 2.6.

Establishments:

For cluster sampling, $\bar{\bar{y}} = 4.49$, $s_b^2 = 20.73$, and $v(\bar{\bar{y}}) = 0.77$.

Further, $s_w^2 = 30.03$, $s^2 = 28.93$, and for simple random sampling, $v(\bar{y}) = 1.07$. Relative precision of cluster sampling is 1.07/.77 = 1.39.

7.5 Forty-six of the 54, that is, 85% of the clusters have more than 10,000 employees.

For the three sample clusters (2, 3, 8), $M_i = (8, 3, 7)$, $c_i = (6, 2, 5)$, and $p_i = (.75, .67, .71)$. Hence $\bar{c} = 13/3$ and $p = 0.71$.

The variance of the c_i is $s^2 = [(6 - 13/3)^2 + (2 - 13/3)^2 + (5 - 13/3)^2]/2 = 13/3$. Now, $v(\bar{c}) = [(9 - 3)/27] \times (13/3) = 26/27$, $v(p) = v(\bar{c})/36 = 0.0267$, and hence S.E.$(p) = 0.1636$.

For the ratio method, from (7.27), $\hat{P} = 13/18 = 0.7222$. From (7.29), $v(\hat{P}) = 250 \times 10^{-6}$ and hence S.E.$(\hat{P}) = 0.0158$.

7.9 With the subscript i = 1, 2, 3, 4 representing the four clusters, Q_i = 0.419, 0.503, 0.503, and 0.575, respectively.

Q_{ij} for (i, j) = (1, 2), (1, 3), (1, 4), (2, 3), (2, 4), and (3, 4), respectively are 0.1292, 0.1292, 0.1607, 0.1667, 0.2071, and 0.2071.

7.11 From (7.37), $v_1(\hat{Y}_{\mathrm{HT}}) = 3134.75$ and S.E.$(\hat{Y}_{\mathrm{HT}}) = 55.99$. From (7.38), $v_2(\hat{Y}_{\mathrm{HT}}) = 4695.22$ and S.E.$(\hat{Y}_{\mathrm{HT}}) = 68.52$.

Since $M_0 = 21$, S.E.s of $\hat{\bar{\bar{Y}}}_{\mathrm{HT}}$ from these two procedures are 55.99/21 = 2.67 and 68.52/21 = 3.26, respectively.

Chapter 8

8.1 For the employments, $S_1^2 = 620.8179$ and $S_2^2 = 5294.77$. Since $N = 9$, $n = 3$, and $m = 3$, from (8.4), $v(\bar{\bar{y}}) = (6/27)S_1^2 + (3/54)S_2^2 = 432.11$.

Since $S^2 = 5057.79$, for the mean of a simple random sample of nine counties, $v(\bar{y}) = [(54 - 9)/54 \times 9](5057.79) = 468.31$.

Relative precision of cluster sampling is 468.31/432.11 = 1.084.

8.3 The first and second terms of (8.9) are 1,266,722.61 and 1,671,459.15. Thus, $V(\hat{Y}) = 2{,}938{,}181.76$ and S.E. $\hat{Y} =$ 1714.11. Hence, S.E. $\bar{\bar{Y}} = 1714.11/54 = 31.74$.

8.5 (a) The first term of (8.12) is equal to 510,148.15 and the second term is the same as in Exercise 8.3. Thus, $V(\hat{Y}_R) = 2{,}1181{,}607.3$ and S.E. $(\hat{Y}_R) = 1{,}477.03$. Hence, S.E.$(\hat{\bar{\bar{Y}}}) = 1477.03/54 = 27.35$.

(b) The above S.E.s for the ratio estimation are about 86% of the corresponding S.E.s in Exercise 8.3.

8.7 From (8.17), $\hat{C} = (9/3)[10/3 + 4/2 + 7(1)] + 37$. The first and second terms of (8.18) are equal to 181.667 and 10.778. Thus, $v(\hat{C}) = 192.45$ and S.E.$(\hat{C}) = 13.87$.

The estimate of the proportion is $\hat{P} = 37/54 = 0.69$ and S.E.$(\hat{P}) = 13.87/(54) = 0.26$.

Since $\bar{p} = 37/66 = 0.56$, from (8.19), $\hat{C}_R = 54(37/66) =$ 30.27. The first term of (8.20) is equal to 140.093 and the second term is equal to 10.778, as above. Thus, $v(\hat{C}_R) =$ 150.871 and S.E.$(\hat{C}_R) = 12.28$. Hence, S.E.$(\bar{p}) = 12.28/54 =$ 0.2274.

The S.E.s of the estimates of the total and proportion with the adjustment for the cluster sizes are about 87% of the S.E.s without the adjustment.

Chapter 9

9.1 (a) $R = 670/550 = 1.2182$. $E(\hat{R}) = 1.2211$, and hence $\hat{R}$ has a bias of 0.0029.

(b) $V(\hat{R}) = 0.004274$ and MSE$(\hat{R}) = 0.004282$. From the approximation in (9.1), $V(\hat{R}) = 0.004225$.

(c) $E[v(\hat{R})] = 0.004136$. The bias of $v(\hat{R})$ for estimating $V(\hat{R})$ in (9.1) and the exact MSE $(\hat{R})$ are respectively equal to -89×10^{-6} and -46×10^{-6}.

9.3 (a) $R = 670/(670 + 550) = 0.5492$. $E(\hat{R}) = 0.5547$, and hence $\hat{R}$ has a bias of 0.0055.

(b) The exact variance and MSE of $\hat{R}$ are equal to 140×10^{-6} and 170×10^{-6}. With an expression similar to (9.1), approximations to the variance and S.E. of $\hat{R}$ are equal to 87×10^{-6} and 0.0093, respectively.

9.5 Denoting heights and weights by (x_i, y_i), $R = \bar{Y}/(\bar{X} + \bar{Y})$. For the five sample units, $(\bar{x}, \bar{y}) = (66, 158.6)$ and $\hat{R} =$ 158.6/224.6 = 0.71.

With $t_i = x_i + y_i$, the sample variance of t_i is $s_t^2 = s_y^2 + s_x^2 + 2s_{xy} = 314.8 + 24 + 2(84.5) = 507.8$, and $s_{yt} = s_{yx} + s_y^2 = 84.5 + 314.8 = 399.3$.

Now, $s_d^2 = s_y^2 + \hat{R}^2 s_t^2 - 2\hat{R}s_{yt} = 3.78$. Following (9.2), $v(\hat{R}) = (1/224.6)^2(10/75)(3.78) = 9.99 \times 10^{-6}$, and hence S.E.$(\hat{R}) = 0.0032$.

The 95% confidence limits for R are given by (.70, .72).

Physicians consider the ideal weight to be 100 pounds for a person five feet tall, and five pounds more for each additional inch of height. For healthy persons over five feet, the above ratio should not be much smaller than one.

9.7 Denoting weight, systolic and diastolic pressures by (x, y_1, y_2), the sample means are $(\bar{x}, \bar{y}_1, \bar{y}_2) = (149, 138, 88.6)$.

Now, $\hat{R}_1 = 138/149 = 0.9262$, $\hat{R}_2 = 88.6/149 = 0.5946$, and $\hat{R}_1 - \hat{R}_2 = 0.3315$. From the sample observations, $s_d^2 = 39.9149$. From (9.7), $v(\hat{R}_1 - \hat{R}_2) = (2/15)(39.9149)/(149)^2 = 0.00024$, and hence S.E.$(\hat{R}_1 - \hat{R}_2) = 0.0155$.

The 95% confidence limits for $(R_1 - R_2)$ are given by (.30, .36).

9.9 With (x, y) denoting the totals for the number of establishments and the employment, $\bar{X} = 18.1$, $\bar{Y} = 265.31$, and $R = 14.66$. Further, $S_x^2 = 94.72$, $S_y^2 = 21396.18$, and $\rho = 0.9429$.

From the above values, for the ratio estimator of the average employment per cluster, $V(\hat{\bar{Y}}_R) = (1 - f) \times (2395.99)/n$. Since $M = 6$, $V(\bar{\bar{Y}}_R) = V(\bar{Y}_R)/36$.

For a simple random sample of n clusters, $V(\bar{y}) = (1 - f) \times (21396.18)/n$ and $V(\bar{\bar{y}}) = V(\bar{y})/36$.

Hence, precision of the ratio estimator relative to the sample mean is equal to $21396.18/2395.99 = 8.93$.

9.11 Denote the public, private, and total enrollments by (x_1, y_1, t_1) for 1990 and by (x_2, y_2, t_2) for 1995.

(a):

	x_1	y_1	x_2	y_2
Mean	352.4	122	356.3	127.4
S.D.	117.82	76.85	121.77	80.03

Mean and S.D. of t_1 are (474.4, 144.19) and of t_2, they are (483.7, 146.66).

Correlations of the public and total enrollments for 1990 and 1995, respectively, are 0.847 0.838. Further, $\mathrm{Cov}(x_1, t_1) = V(x_1) + \mathrm{Cov}(x_1, y_1) = 1483.49$.

Covariances

	x_1	y_1	x_2	y_2
x_1	13,881.0			
y_1	502.6	5,905.3		
x_2	13,799.1	−36.6	14,828.0	
y_2	661.3	6,140.8	136.0	6404.0

Correlations

	x_1	y_1	x_2	y_2
x_1	1			
y_1	.056	1		
x_2	.962	−.004	1	
y_2	.070	.999	.014	1

(b) and (c):
For 1990, $R = 352.4/474.4 = 0.7428$. The ratio estimator for the public enrollments is obtained from $\hat{\bar{X}}_{1R} = (\bar{x}_1/\bar{t}_1) \times (474.4)$. Further, $V(\hat{\bar{X}}_{1R}) = 597.67$ and S.E.$\hat{\bar{X}}_{1R} = 24.45$.

For the sample mean, $V(\bar{x}_1) = 2082.14$, and hence S.E.$(\bar{x}_1) = 45.63$. Relative precision of $\hat{\bar{X}}_{1R}$ is 2082.14/ 597.67 = 3.48.

For 1995, $R = 356.3/483.7 = 0.7366$. The ratio estimator for the public enrollments is obtained from $\hat{\bar{X}}_{2R} = (\bar{x}_2/\bar{t}_2)(483.7)$. Further, $V(\hat{\bar{X}}_{2R}) = 667.67$ and S.E.$(\hat{\bar{X}}_{2R}) = 25.84$.

For the sample mean, $V(\bar{x}_2) = 2224.20$, and hence S.E.$(\bar{x}_2) = 47.16$. Relative precision of $\hat{\bar{X}}_{1R}$ is 2224.20/ 667.67 = 3.33.

9.13

(a):

	x_1	y_1	x_2	y_2
Mean	33.4	7.6	35.4	8.4
S.D.	7.14	5.34	9.26	6.06

Covariances

	x_1	y_1	x_2	y_2
x_1	50.93			
y_1	8.96	28.49		
x_2	60.82	9.07	85.82	
y_2	14.93	31.51	13.93	50.64

Correlations

	x_1	y_1	x_2	y_2
x_1	1			
y_1	.24	1		
x_2	.92	.76	1	
y_2	.35	.78	.25	1

(b) For 1990, $R_1 = 33.4/7.6 = 4.39$, and $V(\hat{R}_1) = 1.35$. For 1995, $R_2 = 35.4/8.4 = 4.21$ and $V(\hat{R}_2) = 1.61$.
(c) $V(\hat{R}_1 - \hat{R}_2) = 0.42$.

9.17 Regression of systolic pressure on weight

Variance of ε_i

	Constant	Prop. to x_i	Prop. to x_i^2
b	0.9082	0.9087	0.9091
Reg. SS	301487.9	1926.5	12.4
$\hat{\sigma}^2$	25.9	0.17	0.00112

For each case, $F_{1,14} = \text{Reg.MS}/\hat{\sigma}^2$ is very large, indicating that the regression is highly significant, that is, the hypothesis that $\beta = 0$ can be easily rejected.

Chapter 10

10.1 After performing (1) through (4), we found that the regression coefficient and the regression estimator for this data are badly biased and have unacceptably large variances and MSEs.

For the data in Table 2.1, we find that $\beta = 0.3254$ and $\rho = 0.41$. Further, the regression and residual MSEs are 3827.04 and 3091.85 with 1 and 4 d.f. Thus $F_{1,4} = 3827.04/3091.85 = 0.81$, which is not significant even at a very large value for the significance level.

Regression through the origin with $V(y_i|x_i)$ to be constant or proportional to x_i and the corresponding estimator for the mean can be suitable for this data; see Exercises 9.1 and 9.2.

10.3 (a) The regression estimators for the means of the systolic and diastolic pressures are $\hat{\bar{Y}}_{1l} = 132.88$ and $\hat{\bar{Y}}_{2l} = 85.72$. Thus, $\hat{\bar{Y}}_{1l} - \hat{\bar{Y}}_{2l} = 47.16$.

(b) $V(\hat{\bar{Y}}_{1l} - \hat{\bar{Y}}_{2l}) = 11.13$ and S.E.$(\hat{\bar{Y}}_{1l} - \hat{\bar{Y}}_{2l}) = 3.34$.

(c) The 95% confidence limits for the difference of the means are (40.61, 53.71).

10.5 (a) For 1990, $V(\hat{\bar{Y}}_l) = 588.39$ and S.E.$(\hat{\bar{Y}}_l) = 24.26$. For 1995, $V(\hat{\bar{Y}}_l) = 662.27$ and S.E.$(\hat{\bar{Y}}_l) = 25.73$.

(b) For both the years, the variances for the regression estimators do not differ much from 597.67 and 667.67 for the ratio estimators found in Exercise 9.11.

10.7 Denoting the public, private, and total enrollments by (y_1, y_2, x), for large n, the difference of the regression estimators for the public and private enrollments can be approximately expressed as $(\hat{\bar{Y}}_{1l} - \hat{\bar{Y}}_{2l}) = (\bar{y}_1 - \bar{y}_2) + (\beta_1 - \beta_2)(\bar{X} - \bar{x})$. Thus, $V(\hat{\bar{Y}}_{1l} - \hat{\bar{Y}}_{2l}) = [(1-f)/n](s_1^2 + s_2^2 + 2s_{12}) + (\beta_1 - \beta_2)^2 s_0^2 + 2(\beta_1 - \beta_2)(s_{10} - s_{20}) = 11.79$. Hence, S.E.$(\hat{\bar{Y}}_{1l} - \hat{\bar{Y}}_{2l} = 3.43$.

Chapter 11

11.1 $p_1 = 150/250 = 0.6$, $v(p_1) = (.6)(.4)/249 = 9.64 \times 10^{-4}$ and hence S.E.$(p_1) = .031$. Thus $p_1 = 60\%$, and it has a S.E. of 3.1%.

The 95% confidence limits are (54, 66)%.

11.3 Denoting husband's and wife's age by h and w, and income by y, we find the following results from the eight completed observations:

Regression with the intercept:

$\hat{y} = -12.35 + 0.45\, w + 1.30\, h$.

The S.E. for the intercept and the two slopes are 25.6, 2.75 and 2.95.

The total, regression and residual *SS* are 1737.8, 920.2, and 817.6 with 7, 2 and 5 d.f., respectively. $F_{2,5} = 2.81$, which is not significant even at the level of 10%.

Regression through the origin:

$\hat{y} = 0.9\, w = 0.495\, h$.

The S.E.s for the two slopes are 2.42 and 2.28.

The total regression and residual *SS* are 1737.9, 882.1, and 855.7 with 8, 2, and 6 d.f., respectively. $F_{2,6} = 3.1$, which is not significant even at the level of 10%.

The final conclusion is that neither of these two multiple regressions with the eight completed observations are preferable to the individual regressions in Example 11.4 of income on husband's age or wife's age.

11.5 As seen in Example 11.5, the proportions of the entire population and the respondents having the attribute of interest are $P = 0.41$ and $P_1 = 0.45$. Further, $W_g = (9/20, 7/20, 4/20)$ and $P_g = (51/90, 28/70, 3/40)$.

The post-stratified estimator is $\hat{P} = \Sigma W_g p_{g1}$ with variance $V(\hat{P}) = \Sigma W_g^2 V(p_{g1})$. For the first type of stratification, $P_{g1} = (36/60, 16/40, 2/20)$ and $\hat{P} = 0.43$. For the second type of stratification, $P_{g1} = (2/8, 18/20, 16/20)$ and $\hat{P} = 0.4625$. Thus, the biases of $\hat{P}$ for these two type of stratification, respectively, are $0.43 - 0.41 = 0.02$ and $0.4625 - 0.41 = 0.0525$.

Responses random:

The sample proportion p_1 obtained from the responding units is unbiased for P. Further, $V(p_1) = [(2000\ 95)/1999](.41)(.59)/95 = 0.002427$, and hence $\text{S.E.}(p_1) = 0.0493$.

For both types of poststratification, $\hat{P}$ is unbiased for P. For the first type, $V(\hat{P}) = 0.002$ and $\text{S.E.}(\hat{P}) = 0.045$. For the second type, $V(\hat{P}) = 0.00226$ and $\text{S.E.}(\hat{P}) = 0.048$.

Responses not random:

In this case, p_1 has a bias of $0.45 - 0.41 = 0.04$. Now $V(p_1) = [(1200 - 95)/1199](.45)(.55)/95 = 0.0024$ and $\text{MSE}(p_1) = 0.004$.

For the first type of stratification, $V(\hat{P}) = 0.00204$ and $\text{MSE}(\hat{P}) = 0.00244$. For the second type of stratification, $V(\hat{P}) = 0.001686$ and $\text{MSE}(\hat{P}) = 0.004443$. The first type of stratification has relatively smaller bias and MSE.

11.7 Known sizes: $N_{12} = 900$ and $N_{22} = 900$.

For the Junior–Senior class, $\bar{\bar{Y}} = (.5)(12.5) + (.5)(14.4) = 13.45$. Further, $v(\hat{\bar{Y}}) = [(900 - 120)/900 \times 120](22) + [(900 - 85)/900 \times 85](32) = 0.0125$, and hence $\text{S.E.}(\hat{\bar{Y}}) = 0.354$.

Estimated sizes: $\hat{N}_{12} = 3600(165/400) = 1485$ and $\hat{N}_{22} = 3600(140/400) = 1260$. Now, $\hat{\bar{Y}} = 13.374$, $v(\hat{\bar{Y}}) = 0.1235$ and $\text{S.E.}(\hat{\bar{Y}}) = 0.351$. The estimate of the mean and its S.E. do not differ much from the above case of known sizes.

Chapter 12

12.1 Deleting the ith observation, $\bar{y}'_i = (n\bar{y} - y_i)/(n - 1)$. The pseudo values are $n\bar{y} - (n - 1)\,\bar{y}'_i = y_i$. Since the jackknife estimator is the average of these values, it is the same as $\bar{y}$.

12.3 For this data, $\bar{x} = 6.51, \bar{y} = 10.48$, $s_x^2 = 72.56$, $s_y^2 = 114.69$, $s_{xy} = 86.896$.

From the above figures, $\hat{R} = 1.61$ and $v(\hat{R}) = 0.0542$.

Since $c_{xx} = 1.71$ and $c_{xy} = 1.27$, $\hat{R}_T = 1.61(1 - .044) = 1.54$.

For the jackknife, we found that $\hat{R}_J = 1.4958$ and $v_J(\hat{R}) = 0.1136$. This estimator for the ratio does not differ much from $\hat{R}$. However, $v(\hat{R})$ and $v_J(\hat{R})$ differ substantially.

12.5 We find from Table 12.1 that except for the sixth unit, the total enrollment (x), is at least five times as much as the private enrollment (y) for the remaining nine units. We have analyzed this problem with these units. Now the mean and variance of x are 174.22 and 13091. For y, the mean and variance are 32.667 and 512. The covariance is $s_{xy} = 2273$.

(a) The regression coefficient is $b = 2273/13091 = 0.1736$. Further, $\hat{\sigma}^2 = 134.26$ and $v(b) = 134.26/104728 = 0.00128$.

(b) For the jackknife, $b_J = 0.1746$ and $v(b_J) = 0.0044$.

This estimate for the regression coefficient does not differ much from the classical estimator b, but there is much difference between the variances for b obtained from these two procedures.

12.7 Denoting the numerator of $V(\hat{R})$ by V, the covariance between the pseudo values can be expressed as Cov $(\hat{R}_i^{\prime}, \hat{R}_j^{\prime}) = n^2 V/n - 2n(n-1)V/(n-1) + (n-1)^2\ V/(n-2) = V/(n-2) = V/(n-2)$, which vanishes for large n.