Chapter 8: Estimation

8.1 Let $B = B(\hat{\theta})$. Then,

$$MSE(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right] = E\left[(\hat{\theta} - E(\hat{\theta}) + B)^2\right] = E\left[\left(\hat{\theta} - E(\hat{\theta})\right)^2\right] + E(B^2) + 2B \times E\left[\hat{\theta} - E(\hat{\theta})\right]$$

$$= V(\hat{\theta}) + B^2.$$

8.2 **a.** The estimator $\hat{\theta}$ is unbiased if $E(\hat{\theta}) = \theta$. Thus, $B(\hat{\theta}) = 0$.
b. $E(\hat{\theta}) = \theta + 5$.

8.3 **a.** Using Definition 8.3, $B(\hat{\theta}) = a\theta + b - \theta = (a - 1)\theta + b$.
b. Let $\hat{\theta}^* = (\hat{\theta} - b)/a$.

8.4 **a.** They are equal.
b. $MSE(\hat{\theta}) > V(\hat{\theta})$.

8.5 **a.** Note that $E(\hat{\theta}^*) = \theta$ and $V(\hat{\theta}^*) = V[(\hat{\theta} - b)/a] = V(\hat{\theta})/a^2$. Then,

$$\text{MSE}(\hat{\theta}^*) = V(\hat{\theta}^*) = V(\hat{\theta})/a^2.$$

b. Note that $\text{MSE}(\hat{\theta}) = V(\hat{\theta}) + B(\hat{\theta}) = V(\hat{\theta}) + [(a-1)\theta + b]^2$. A sufficiently large value of a will force $\text{MSE}(\hat{\theta}^*) < \text{MSE}(\hat{\theta})$. Example: $a = 10$.

c. A amply small value of a will make $\text{MSE}(\hat{\theta}^*) > \text{MSE}(\hat{\theta})$. Example: $a = .5$, $b = 0$.

8.6 **a.** $E(\hat{\theta}_3) = aE(\hat{\theta}_1) + (1-a)E(\hat{\theta}_2) = a\theta + (1-a)\theta = \theta$.

b. $V(\hat{\theta}_3) = a^2V(\hat{\theta}_1) + (1-a)^2V(\hat{\theta}_2) = a^2\sigma_1^2 + (1-a)\sigma_2^2$, since it was assumed that $\hat{\theta}_1$ and $\hat{\theta}_2$ are independent. To minimize $V(\hat{\theta}_3)$, we can take the first derivative (with respect to a), set it equal to zero, to find

$$a = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}.$$

(One should verify that the second derivative test shows that this is indeed a minimum.)

8.7 Following Ex. 8.6 but with the condition that $\hat{\theta}_1$ and $\hat{\theta}_2$ are not independent, we find

$$V(\hat{\theta}_3) = a^2\sigma_1^2 + (1-a)\sigma_2^2 + 2a(1-a)c.$$

Using the same method w/ derivatives, the minimum is found to be

$$a = \frac{\sigma_2^2 - c}{\sigma_1^2 + \sigma_2^2 - 2c}.$$

8.8 **a.** Note that $\hat{\theta}_1$, $\hat{\theta}_2$, $\hat{\theta}_3$ and $\hat{\theta}_5$ are simple linear combinations of Y_1, Y_2, and Y_3. So, it is easily shown that all four of these estimators are *unbiased*. From Ex. 6.81 it was shown that $\hat{\theta}_4$ has an exponential distribution with mean $\theta/3$, so this estimator is biased.

b. It is easily shown that $V(\hat{\theta}_1) = \theta^2$, $V(\hat{\theta}_2) = \theta^2/2$, $V(\hat{\theta}_3) = 5\theta^2/9$, and $V(\hat{\theta}_5) = \theta^2/9$, so the estimator $\hat{\theta}_5$ is unbiased and has the smallest variance.

8.9 The density is in the form of the exponential with mean $\theta + 1$. We know that $\overline{Y}$ is unbiased for the mean $\theta + 1$, so an unbiased estimator for θ is simply $\overline{Y} - 1$.

8.10 **a.** For the Poisson distribution, $E(Y) = \lambda$ and so for the random sample, $E(\overline{Y}) = \lambda$. Thus, the estimator $\hat{\lambda} = \overline{Y}$ is unbiased.

b. The result follows from $E(Y) = \lambda$ and $E(Y^2) = V(Y) + \lambda^2 = 2\lambda^2$, so $E(C) = 4\lambda + \lambda^2$.

c. Since $E(\overline{Y}) = \lambda$ and $E(\overline{Y}^2) = V(\overline{Y}) + [E(\overline{Y})]^2 = \lambda^2/n + \lambda^2 = \lambda^2(1 + 1/n)$. Then, we can construct an unbiased estimator $\hat{\theta} = \overline{Y}^2 + \overline{Y}(4 - 1/n)$.

8.11 The third central moment is defined as

$$E[(Y-\mu)^3] = E[(Y-3)^3] = E(Y^3) - 9E(Y^2) + 54 .$$

Using the unbiased estimates $\hat{\theta}_2$ and $\hat{\theta}_3$, it can easily be shown that $\hat{\theta}_3 - 9\hat{\theta}_2 + 54$ is an unbiased estimator.

8.12 **a.** For the uniform distribution given here, $E(Y_i) = \theta + .5$. Hence, $E(\overline{Y}) = \theta + .5$ so that $B(\overline{Y}) = .5$.

b. Based on $\overline{Y}$, the unbiased estimator is $\overline{Y} - .5$.

c. Note that $V(\overline{Y}) = 1/(12n)$ so $\text{MSE}(\overline{Y}) = 1/(12n) + .25$.

8.13 **a.** For a random variable Y with the binomial distribution, $E(Y) = np$ and $V(Y) = npq$, so $E(Y^2) = npq + (np)^2$. Thus,

$$E\left\{n\left(\tfrac{Y}{n}\right)\left[1 - \tfrac{Y}{n}\right]\right\} = E(Y) - \tfrac{1}{n}E(Y^2) = np - pq - np^2 = (n-1)pq .$$

b. The unbiased estimator should have expected value npq, so consider the estimator

$$\hat{\theta} = \left(\tfrac{n}{n-1}\right)n\left(\tfrac{Y}{n}\right)\left[1 - \tfrac{Y}{n}\right].$$

8.14 Using standard techniques, it can be shown that $E(Y) = \left(\frac{\alpha}{\alpha+1}\right)\theta$, $E(Y^2) = \left(\frac{\alpha}{\alpha+2}\right)\theta^2$. Also, it is easily shown that $Y_{(n)}$ follows the power family with parameters $n\alpha$ and θ.

a. From the above, $E(\hat{\theta}) = E(Y_{(n)}) = \left(\frac{n\alpha}{n\alpha+1}\right)\theta$, so that the estimator is biased.

b. Since α is known, the unbiased estimator is $\left(\frac{n\alpha+1}{n\alpha}\right)\hat{\theta} = \left(\frac{n\alpha+1}{n\alpha}\right)Y_{(n)}$.

c. $MSE(Y_{(n)}) = E[(Y_{(n)} - \theta)^2] = E(Y_{(n)}^2) - 2\theta E(Y_{(n)}) + \theta^2 = \frac{2}{(n\alpha+1)(n\alpha+2)}\theta^2$.

8.15 Using standard techniques, it can be shown that $E(Y) = (3/2)\beta$, $E(Y^2) = 3\beta^2$. Also it is easiliy shown that $Y_{(1)}$ follows the Pareto family with density function

$$g_{(1)}(y) = 3n\beta^{3n}y^{-(3n+1)},\ y \geq \beta.$$

Thus, $E(Y_{(1)}) = \left(\frac{3n}{3n-1}\right)\beta$ and $E(Y_{(1)}^2) = \frac{3n}{3n-2}\beta^2$.

a. With $\hat{\beta} = Y_{(1)}$, $B(\hat{\beta}) = \left(\frac{3n}{3n-1}\right)\beta - \beta = \left(\frac{1}{3n-1}\right)\beta$.

b. Using the above, $MSE(\hat{\beta}) = MSE(Y_{(1)}) = E(Y_{(1)}^2) - 2\beta E(Y_{(1)}) + \beta^2 = \frac{2}{(3n-1)(3n-2)}\beta^2$.

8.16 It is known that $(n-1)S^2/\sigma^2$ is chi–square with n–1 degrees of freedom.

a. $E(S) = E\left\{\frac{\sigma}{\sqrt{n-1}}\left[\frac{(n-1)S^2}{\sigma^2}\right]^{1/2}\right\} = \frac{\sigma}{\sqrt{n-1}}\int_0^\infty v^{1/2}\frac{1}{\Gamma[(n-1)/2]2^{(n-1)/2}}v^{(n-1)/2}e^{-v/2}dv = \frac{\sigma}{\sqrt{n-1}}\frac{\sqrt{2}\Gamma(n/2)}{\Gamma[(n-1)/2]}$.

b. The estimator $\hat{\sigma} = \frac{\sqrt{n-1}\Gamma[(n-1)/2]}{\sqrt{2}\Gamma(n/2)}S$ is unbiased for σ.

c. Since $E(\overline{Y}) = \mu$, the unbiased estimator of the quantity is $\overline{Y} - z_\alpha\hat{\sigma}$.

8.17 It is given that $\hat{p}_1$ is unbiased, and since $E(Y) = np$, $E(\hat{p}_2) = (np + 1)/(n+2)$.

a. $B(\hat{p}_2) = (np + 1)/(n+2) - p = (1-2p)/(n+2)$.

b. Since $\hat{p}_1$ is unbiased, $\text{MSE}(\hat{p}_1) = \text{V}(\hat{p}_1) = p(1-p)/n$. $\text{MSE}(\hat{p}_2) = V(\hat{p}_2) + B(\hat{p}_2) = \frac{np(1-p)+(1-2p)^2}{(n+2)^2}$.

c. Considering the inequality

$$\frac{np(1-p)+(1-2p)^2}{(n+2)^2} < \frac{p(1-p)}{n},$$

this can be written as

$$(8n+4)p^2 - (8n+4)p + n < 0.$$

Solving for p using the quadratic formula, we have

$$p = \frac{8n+4 \pm \sqrt{(8n+4)^2 - 4(8n+4)n}}{2(8n+4)} = \frac{1}{2} \pm \sqrt{\frac{n+1}{8n+4}}.$$

So, p will be close to .5.

8.18 Using standard techniques from Chapter 6, is can be shown that the density function for $Y_{(1)}$ is given by

$$g_{(1)}(y) = \frac{n}{\theta}\left(1 - \frac{y}{\theta}\right)^{n-1},\ 0 \leq y \leq \theta.$$

So, $E(Y_{(1)}) = \frac{\theta}{n+1}$ and so an unbiased estimator for θ is $(n+1)Y_{(1)}$.

8.19 From the hint, we know that $E(Y_{(1)}) = \beta/n$ so that $\hat{\theta} = nY_{(1)}$ is unbiased for β. Then, $\text{MSE}(\hat{\theta}) = V(\hat{\theta}) + B(\hat{\theta}) = V(nY_{(1)}) = n^2V(Y_{(1)}) = \beta^2$.

8.20 If Y has an exponential distribution with mean θ, then by Ex. 4.11, $E(\sqrt{Y}) = \sqrt{\pi\theta}/2$.

a. Since Y_1 and Y_2 are independent, $E(X) = \pi\theta/4$ so that $(4/\pi)X$ is unbiased for θ.

b. Following part a, it is easily seen that $E(W) = \pi^2\theta^2/16$, so $(4^2/\pi^2)W$ is unbiased for θ^2.

8.21 Using Table 8.1, we can estimate the population mean by $\bar{y} = 11.5$ and use a two–standard–error bound of $2(3.5)/\sqrt{50} = .99$. Thus, we have $11.5 \pm .99$.

8.22 (Similar to Ex. 8.21) The point estimate is $\bar{y} = 7.2\%$ and a bound on the error of estimation is $2(5.6)/\sqrt{200} = .79\%$.

8.23 **a.** The point estimate is $\bar{y} = 11.3$ ppm and an error bound is $2(16.6)/\sqrt{467} = 1.54$ ppm.

b. The point estimate is 46.4 – 45.1 = 1.3 and an error bound is $2\sqrt{\frac{(9.8)^2}{191} + \frac{(10.2)^2}{467}} = 1.7$.

c. The point estimate is .78 – .61 = .17 and an error bound is $2\sqrt{\frac{(.78)(.22)}{467} + \frac{(.61)(.39)}{191}} = .08$.

8.24 Note that by using a two–standard–error bound, $2\sqrt{\frac{(.69)(.31)}{1001}} = .0292 \approx .03$. Constructing this as an interval, this is (.66, .72). We can say that there is little doubt that the true (population) proportion falls in this interval. Note that the value 50% is far from the interval, so it is clear that a majority did feel that the cost of gasoline was a problem.

8.25 We estimate the difference to be 2.4 – 3.1 = –.7 with an error bound of $2\sqrt{\frac{1.44+2.64}{100}} = .404$.

8.26 **a.** The estimate of the true population proportion who think humans should be sent to Mars is .49 with an error bound of $2\sqrt{\frac{.49(.51)}{1093}} = .03$.

b. The standard error is given by $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, and this is maximized when $\hat{p} = .5$. So, a conservative error bound that could be used for all sample proportions (with $n = 1093$) is $2\sqrt{\frac{.5(.5)}{1093}} = .0302$ (or 3% as in the above).

8.27 **a.** The estimate of p is the sample proportion: 592/985 = .601, and an error bound is given by $2\sqrt{\frac{.601(.399)}{985}} = .031$.

b. The above can be expressed as the interval (.570, .632). Since this represents a clear majority for the candidate, it appears certain that the republican will be elected. Following Example 8.2, we can be reasonably confident by this statement.

c. The group of "likely voters" is not necessarily the same as "definite voters."

8.28 The point estimate is given by the difference of the sample proportions: .70 – .54 = .16 and an error bound is $2\sqrt{\frac{.7(.3)}{180}+\frac{.54(.46)}{100}}$ = .121.

8.29 **a.** The point estimate is the difference of the sample proportions: .45 – .51 = –.06, and an error bound is $2\sqrt{\frac{.45(.55)}{1001}+\frac{.51(.49)}{1001}}$ = .045.

b. The above can be expressed as the interval (–.06 – .045, –.06 + .045) or (–.105, –.015). Since the value 0 is not contained in the interval, it seems reasonable to claim that fan support for baseball is greater at the end of the season.

8.30 The point estimate is .45 and an error bound is $2\sqrt{\frac{.45(.55)}{1001}}$ = .031. Since 10% is roughly three times the two–standard–error bound, it is not likely (assuming the sample was indeed a randomly selected sample).

8.31 **a.** The point estimate is the difference of the sample proportions: .93 – .96 = –.03, and an error bound is $2\sqrt{\frac{.93(.07)}{200}+\frac{.96(.04)}{450}}$ = .041.

b. The above can be expressed as the interval (–.071, .011). Note that the value zero is contained in the interval, so there is reason to believe that the two pain relievers offer the same relief potential.

8.32 With n = 20, the sample mean amount $\bar{y}$ = 197.1 and the standard deviation s = 90.86.

- The *total accounts receivable* is estimated to be 500($\bar{y}$) = 500(197.1) = 98,550. The standard deviation of this estimate is found by $\sqrt{V(500\bar{Y})} = 500\frac{\sigma}{\sqrt{20}}$. So, this can be estimated by 500(90.86)/$\sqrt{20}$ = 10158.45 and an error bound is given by 2(10158.46) = 20316.9.

- With $\bar{y}$ = 197.1, an error bound is 2(90.86)/$\sqrt{20}$ = 40.63. Expressed as an interval, this is (197.1 – 40.63, 197.1 + 40.63) or (156.47, 237.73). So, it is unlikely that the average amount exceeds $250.

8.33 The point estimate is 6/20 = .3 and an error bound is $2\sqrt{\frac{.3(.7)}{20}}$ = .205. If 80% comply, and 20% fail to comply. This value lies within our error bound of the point estimate, so it is likely.

8.34 An unbiased estimator of λ is $\bar{Y}$, and since $\sqrt{V(\bar{Y})} = \sqrt{\lambda / n}$, an unbiased estimator of the standard error of is $\sqrt{\bar{Y} / n}$.

8.35 Using the result of Ex. 8.34:

a. The point estimate is $\bar{y}$ = 20 and a bound on the error of estimation is $2\sqrt{20/50}$ = 1.265.

b. The point estimate is the difference of the sample mean: 20 – 23 = –3.

8.36 An unbiased estimator of θ is $\overline{Y}$, and since $\sqrt{V(\overline{Y})} = \theta/\sqrt{n}$, an unbiased estimator of the standard error of is $\overline{Y}/\sqrt{n}$.

8.37 Refer to Ex. 8.36: with n = 10, an estimate of θ = $\overline{y}$ = 1020 and an error bound is $2\left(1000/\sqrt{10}\right)$ = 645.1.

8.38 To find an unbiased estimator of $V(Y) = \frac{1}{p^2} - \frac{1}{p}$, note that $E(Y) = \frac{1}{p}$ so Y is an unbiased estimator of $\frac{1}{p}$. Further, $E(Y^2) = V(Y) + [E(Y)]^2 = \frac{2}{p^2} - \frac{1}{p}$ so $E(Y^2 + Y) = \frac{2}{p^2}$. Therefore, an unbiased estimate of $V(Y)$ is $\frac{Y^2+Y}{2} + Y = \frac{Y^2-Y}{2}$.

8.39 Using Table 6 with 4 degrees of freedom, $P\left(.71072 \le 2Y/\beta \le 9.48773\right)$ = .90. So,

$$P\left(\tfrac{2Y}{9.48773} \le \beta \le \tfrac{2Y}{.71072}\right) = .90$$

and $\left(\frac{2Y}{9.48773}, \frac{2Y}{.71072}\right)$ forms a 90% CI for β.

8.40 Use the fact that $Z = \frac{Y-\mu}{\sigma}$ has a standard normal distribution. With σ = 1:

a. The 95% CI is (Y – 1.96, Y + 1.96) since

$$P(-1.96 \le Y - \mu \le 1.96) = P(Y - 1.96 \le \mu \le Y + 1.96) = .95.$$

b. The value Y + 1.645 is the 95% upper limit for μ since

$$P(Y - \mu \le 1.645) = P(\mu \le Y + 1.645) = .95.$$

c. Similarly, Y – 1.645 is the 95% lower limit for μ.

8.41 Using Table 6 with 1 degree of freedom:

a. $.95 = P(.0009821 \le Y^2/\sigma^2 \le 5.02389) = P(Y^2/5.02389 \le \sigma^2 \le Y^2/.0009821)$.

b. $.95 = P(.0039321 \le Y^2/\sigma^2) = P(\sigma^2 \le Y^2/.0039321)$.

c. $.95 = P(Y^2/\sigma^2 \le 3.84146) = P(Y^2/3.84146 \le \sigma^2)$.

8.42 Using the results from Ex. 8.41, the square–roots of the boundaries can be taken to obtain interval estimates σ:

a. $Y/2.24 \le \sigma \le Y/.0313$.

b. $\sigma \le Y/.0627$.

c. $\sigma \ge Y/1.96$.

8.43 **a.** The distribution function for $Y_{(n)}$ is $G_n(y) = \left(\frac{y}{\theta}\right)^n$, $0 \le y \le \theta$, so the distribution function for U is given by

$$F_U(u) = P(U \le u) = P(Y_{(n)} \le \theta u) = G_n(\theta u) = u,\ 0 \le y \le 1.$$

b. (Similar to Example 8.5) We require the value a such that $P\left(\frac{Y_{(n)}}{\theta} \le a\right) = F_U(a) = .95$.

Therefore, $a^n = .95$ so that $a = (.95)^{1/n}$ and the lower confidence bound is $[Y_{(n)}](.95)^{-1/n}$.

8.44 **a.** $F_Y(y) = P(Y \le y) = \int_0^y \frac{2(\theta - t)}{\theta^2} dt = \frac{2y}{\theta} - \frac{y^2}{\theta^2}$, $0 < y < \theta$.

b. The distribution of $U = Y/\theta$ is given by
$F_U(u) = P(U \le u) = P(Y \le \theta u) = F_Y(\theta u) = 2u - u^2 = 2u(1-u)$, $0 < u < 1$. Since this distribution does not depend on θ, $U = Y/\theta$ is a pivotal quantity.

c. Set $P(U \le a) = F_Y(a) = 2a(1 - a) = .9$ so that the quadratic expression is solved at $a = 1 - \sqrt{.10} = .6838$ and then the 90% lower bound for θ is $Y/.6838$.

8.45 Following Ex. 8.44, set $P(U \ge b) = 1 - F_Y(b) = 1 - 2b(1 - b) = .9$, thus $b = 1 - \sqrt{.9} = .05132$ and then the 90% upper bound for θ is $Y/.05132$.

8.46 Let $U = 2Y/\theta$ and let $m_Y(t)$ denote the mgf for the exponential distribution with mean θ. Then:

a. $m_U(t) = E(e^{tU}) = E(e^{t2Y/\theta}) = m_Y(2t/\theta) = (1-2t)^{-1}$. This is the mgf for the chi–square distribution with one degree of freedom. Thus, U has this distribution, and since the distribution does not depend on θ, U is a pivotal quantity.

b. Using Table 6 with 2 degrees of freedom, we have
$$P(.102587 \le 2Y/\theta \le 5.99147) = .90.$$
So, $\left(\frac{2Y}{5.99147}, \frac{2Y}{.102587}\right)$ represents a 90% CI for θ.

c. They are equivalent.

8.47 Note that for all i, the mgf for Y_i is $m_Y(t) = (1-\theta t)^{-1}$, $t < 1/\theta$.

a. Let $U = 2\sum_{i=1}^n Y_i/\theta$. The mgf for U is
$$m_U(t) = E(e^{tU}) = [m_Y(2t/\theta)]^n = (1-2t)^{-n},\ t < 1/2.$$
This is the mgf for the chi–square distribution with $2n$ degrees of freedom. Thus, U has this distribution, and since the distribution does not depend on θ, U is a pivotal quantity.

b. Similar to part b in Ex. 8.46, let $\chi^2_{.975}, \chi^2_{.025}$ be percentage points from the chi–square distribution with $2n$ degrees of freedom such that
$$P\left(\chi^2_{.975} \le 2\sum_{i=1}^n Y_i/\theta \le \chi^2_{.025}\right) = .95.$$

So, $\left(\frac{2\sum_{i=1}^{n} Y_i}{\chi^2_{.975}}, \frac{2\sum_{i=1}^{n} Y_i}{\chi^2_{.025}}\right)$ represents a 95% CI for θ.

c. The CI is $\left(\frac{2(7)(4.77)}{26.1190}, \frac{2(7)(4.77)}{5.62872}\right)$ or (2.557, 11.864).

8.48 (Similar to Ex. 8.47) Note that for all i, the mgf for Y_i is $m_Y(t) = (1-\beta)^{-2}$, $t < 1/\beta$.

a. Let $U = 2\sum_{i=1}^{n} Y_i / \beta$. The mgf for U is

$$m_U(t) = E(e^{tU}) = \left[m_Y(2t/\beta)\right]^n = (1-2t)^{-2n},\ t < 1/2.$$

This is the mgf for the chi–square distribution with 4n degrees of freedom. Thus, U has this distribution, and since the distribution does not depend on θ, U is a pivotal quantity.

b. Similar to part b in Ex. 8.46, let $\chi^2_{.975}$, $\chi^2_{.025}$ be percentage points from the chi–square distribution with 4n degrees of freedom such that

$$P\left(\chi^2_{.975} \le 2\sum_{i=1}^{n} Y_i / \beta \le \chi^2_{.025}\right) = .95.$$

So, $\left(\frac{2\sum_{i=1}^{n} Y_i}{\chi^2_{.975}}, \frac{2\sum_{i=1}^{n} Y_i}{\chi^2_{.025}}\right)$ represents a 95% CI for β.

c. The CI is $\left(\frac{2(5)(5.39)}{34.1696}, \frac{2(5)(5.39)}{9.59083}\right)$ or (1.577, 5.620).

8.49 **a.** If α = m (a known integer), then $U = 2\sum_{i=1}^{n} Y_i / \beta$ still a pivotal quantity and using a mgf approach it can be shown that U has a chi–square distribution with mn degrees of freedom. So, the interval is

$$\left(\frac{2\sum_{i=1}^{n} Y_i}{\chi^2_{1-\alpha/2}}, \frac{2\sum_{i=1}^{n} Y_i}{\chi^2_{\alpha/2}}\right),$$

where $\chi^2_{1-\alpha/2}$, $\chi^2_{\alpha/2}$ are percentage points from the chi–square distribution with mn degrees of freedom.

b. The quantity $U = \sum_{i=1}^{n} Y_i / \beta$ is distributed as gamma with shape parameter cn and scale parameter 1. Since c is known, percentiles from this distribution can be calculated from this gamma distribution (denote these as $\gamma_{1-\alpha/2}$, $\gamma_{\alpha/2}$) so that similar to part a, the CI is

$$\left(\frac{\sum_{i=1}^{n} Y_i}{\gamma_{1-\alpha/2}}, \frac{\sum_{i=1}^{n} Y_i}{\gamma^2_{\alpha/2}}\right).$$

c. Following the notation in part b above, we generate the percentiles using the Applet:

$$\chi_{.975} = 16.74205,\ \chi_{.025} = 36.54688$$

Thus, the CI is $\left(\frac{10(11.36)}{36.54688}, \frac{10(11.36)}{16.74205}\right)$ or (3.108, 6.785).

8.50 **a.** –.1451
b. .2251
c. Brand *A* has the larger proportion of failures, 22.51% greater than Brand *B*.
d. Brand *B* has the larger proportion of failures, 14.51% greater than Brand *A*.
e. There is no evidence that the brands have different proportions of failures, since we are not confident that the brand difference is strictly positive or negative.

8.51 **a.-f**. Answers vary.

8.52 **a.-c**. Answers vary.
d. The proportion of intervals that capture p should be close to .95 (the confidence level).

8.53 **a.** i. Answers vary. ii. smaller confidence level, larger sample size, smaller value of p.
b. Answers vary.

8.54 **a.** The interval is not calculated because the length is zero (the standard error is zero).
b.-d. Answers vary.
e. The sample size is not large (consider the validity of the normal approximation to the binomial).

8.55 Answers vary, but with this sample size, a normal approximation is appropriate.

8.56 **a.** With $z_{.01}$ = 2.326, the 98% CI is $.45 \pm 2.326\sqrt{\frac{.45(.55)}{800}}$ or .45 ± .041.
b. Since the value .50 is not contained in the interval, there is not compelling evidence that a majority of adults feel that movies are getting better.

8.57 With $z_{.005}$ = 2.576, the 99% interval is $.51 \pm 2.576\sqrt{\frac{.51(.49)}{1001}}$ or .51 ± .04. We are 99% confident that between 47% and 55% of adults in November, 2003 are baseball fans.

8.58 The parameter of interest is μ = mean number of days required for treatment. The 95% CI is approximately $\bar{y} \pm z_{.025}\left(s/\sqrt{n}\right)$, or $5.4 \pm 1.96(3.1/\sqrt{500})$ or (5.13, 5.67).

8.59 **a.** With $z_{.05}$ = 1.645, the 90% interval is $.78 \pm 1.645\sqrt{\frac{.78(.22)}{1030}}$ or .78 ± .021.
b. The lower endpoint of the interval is .78 – .021 = .759, so there is evidence that the true proportion is greater than 75%.

8.60 **a.** With $z_{.005}$ = 2.576, the 99% interval is $98.25 \pm 2.576\left(.73/\sqrt{130}\right)$ or 98.25 ± .165.

b. Written as an interval, the above is (98.085, 98.415). So, the "normal" body temperature measurement of 98.6 degrees is not contained in the interval. It is possible that the standard for "normal" is no longer valid.

8.61 With $z_{.025}$ = 1.96, the 95% CI is $167.1 - 140.9 \pm 1.96\sqrt{\frac{(24.3)^2+(17.6)^2}{30}}$ or (15.46, 36.94).

8.62 With $z_{.005}$ = 2.576, the approximate 99% CI is $24.8 - 21.3 \pm 2.576\sqrt{\frac{(7.1)^2}{34} + \frac{(8.1)^2}{41}}$ or $(-1.02, 8.02)$. With 99% confidence, the difference in mean molt time for normal males versus those split from their mates is between (–1.02, 8.02).

8.63 **a.** With $z_{.025}$ = 1.96, the 95% interval is $.78 \pm 1.96\sqrt{\frac{.78(.22)}{1000}}$ or .78 ± .026 or (.754, .806).

b. The margin of error reported in the article is larger than the 2.6% calculated above. Assuming that a 95% CI was calculated, a value of p = .5 gives the margin of error 3.1%.

8.64 **a.** The point estimates are .35 (sample proportion of 18-34 year olds who consider themselves patriotic) and .77 (sample proportion of 60+ year olds who consider themselves patriotic. So, a 98% CI is given by (here, $z_{.01}$ = 2.326)

$$.77 - .35 \pm 2.326\sqrt{\frac{(.77)(.23)}{150} + \frac{(.35)(.65)}{340}}$$ or .42 ± .10 or (.32, .52).

b. Since the value for the difference .6 is outside of the above CI, this is not a likely value.

8.65 **a.** The 98% CI is, with $z_{.01}$ = 2.326, is

$$.18 - .12 \pm 2.326\sqrt{\frac{.18(.82)+.12(.88)}{100}}$$ or .06 ± .117 or (–.057, .177).

b. Since the interval contains both positive and negative values, it is likely that the two assembly lines produce the same proportion of defectives.

8.66 **a.** With $z_{.05}$ = 1.645, the 90% CI for the mean posttest score for all BACC students is $18.5 \pm 1.645\left(\frac{8.03}{\sqrt{365}}\right)$ or 18.5 ± .82 or (17.68, 19.32).

b. With $z_{.025}$ = 1.96, the 95% CI for the difference in the mean posttest scores for BACC and traditionally taught students is $(18.5 - 16.5) \pm 1.96\sqrt{\frac{(8.03)^2}{365} + \frac{(6.96)^2}{298}}$ or 2.0 ± 1.14.

c. Since 0 is outside of the interval, there is evidence that the mean posttest scores are different.

8.67 **a.** The 95% CI is $7.2 \pm 1.96\sqrt{\frac{8.8}{60}}$ or 7.2 ± .75 or (6.45, 7.95).

b. The 90% CI for the difference in the mean densities is $(7.2-4.7)\pm 1.645\sqrt{\frac{8.8}{60}+\frac{4.9}{90}}$ or 2.5 ± .74 or (1.76, 3.24).

c. Presumably, the population is ship sightings for all summer and winter months. It is quite possible that the days used in the sample were not randomly selected (the months were chosen in the same year.)

8.68 **a.** Recall that for the multinomial, $V(Y_i) = np_iq_i$ and $\text{Cov}(Y_i,Y_j) = -np_ip_j$ for $i \neq j$. Hence,

$$V(Y_1 - Y_2) = V(Y_1) + V(Y_2) - 2\text{Cov}(Y_1, Y_2) = np_1q_1 + np_2q_2 + 2np_1p_2.$$

b. Since $\hat{p}_1 - \hat{p}_2 = \frac{Y_1 - Y_2}{n}$, using the result in part a we have

$$V(\hat{p}_1 - \hat{p}_2) = \tfrac{1}{n}\left(p_1q_1 + p_2q_2 + 2p_1p_2\right).$$

Thus, an approximate 95% CI is given by

$$\hat{p}_1 - \hat{p}_2 \pm 1.96\sqrt{\tfrac{1}{n}\left(\hat{p}_1\hat{q}_1 + \hat{p}_2\hat{q}_2 + 2\hat{p}_1\hat{p}_2\right)}$$

Using the supplied data, this is

$.06 - .16 \pm 1.96\sqrt{\frac{1}{500}\left(.06(.94) + .16(.84) + 2(.06)(.16)\right)} = -.10 \pm .04$ or (–.14, –.06).

8.69 For the independent counts Y_1, Y_2, Y_3, and Y_4, the sample proportions are $\hat{p}_i = Y_i / n_i$ and $V(\hat{p}_i) = p_iq_i / n_i$ for $i = 1, 2, 3, 4$. The interval of interest can be constructed as

$$(\hat{p}_3 - \hat{p}_1) - (\hat{p}_4 - \hat{p}_2) \pm 1.96\sqrt{V[(\hat{p}_3 - \hat{p}_1) - (\hat{p}_4 - \hat{p}_2)]}\,.$$

By independence, this is

$$(\hat{p}_3 - \hat{p}_1) - (\hat{p}_4 - \hat{p}_2) \pm 1.96\sqrt{\tfrac{1}{n}[\hat{p}_3\hat{q}_3 + \hat{p}_1\hat{q}_1 + \hat{p}_4\hat{q}_4 + \hat{p}_2\hat{q}_2]}\,.$$

Using the sample data, this is

$$(.69 - .65) - (.25 - .43) \pm 1.96\sqrt{\tfrac{1}{500}[.65(.35) + .43(.57) + .69(.31) + .25(.75)}$$

or .22 ± .34 or (–.12, .56)

8.70 As with Example 8.9, we must solve the equation $1.96\sqrt{\frac{pq}{n}} = B$ for n.

a. With $p = .9$ and $B = .05$, $n = 139$.
b. If p is unknown, use $p = .5$ so $n = 385$.

8.71 With $B = 2$, $\sigma = 10$, $n = 4\sigma^2/B^2$, so $n = 100$.

8.72 **a.** Since the true proportions are unknown, use .5 for both to compute an error bound (here, we will use a multiple of 1.96 that correlates to a 95% CI):

$$1.96\sqrt{\frac{.5(.5)}{1000} + \frac{.5(.5)}{1000}} = .044.$$

b. Assuming that the two sample sizes are equal, solve the relation

$$1.645\sqrt{\frac{.5(.5)}{n} + \frac{.5(.5)}{n}} = .02\,,$$

so $n = 3383$.

8.73 From the previous sample, the proportion of 'tweens who understand and enjoy ads that are silly in nature is .78. Using this as an estimate of p, we estimate the sample size as

$$2.576\sqrt{\tfrac{.78(.22)}{n}} = .02 \text{ or } n = 2847.$$

8.74 With $B = .1$ and $\sigma = .5$, $n = (1.96)^2\sigma^2/B^2$, so $n = 97$. If all of the specimens were selected from a single rainfall, the observations would not be independent.

8.75 Here, $1.645\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = .1$, but $\sigma_1^2 = \sigma_2^2 = .25$, $n_1 = n_2 = n$, so sample $n = 136$ from each location.

8.76 For $n_1 = n_2 = n$ and by using the estimates of population variances given in Ex. 8.61, we can solve $1.645\sqrt{\frac{(24.3)^2+(17.6)^2}{n}} = 5$ so that $n = 98$ adults must be selected from each region.

8.77 Using the estimates $\hat{p}_1 = .7$, $\hat{p}_2 = .54$, the relation is $1.645\sqrt{\frac{.7(.3)+.54(.46)}{n}} = .05$ so $n = 497$.

8.78 Here, we will use the estimates of the true proportions of defectives from Ex. 8.65. So, with a bound $B = (.2)/2 = .1$, the relation is $1.96\sqrt{\frac{.18(.82)+.12(.88)}{n}} = .1$ so $n = 98$.

8.79 **a.** Here, we will use the estimates of the population variances for the two groups of students:

$$2.576\sqrt{\frac{(8.03)^2}{n} + \frac{(6.96)^2}{n}} = .5,$$

so $n = 2998$ students from each group should be sampled.

b. For comparing the mean pretest scores, $s_1 = 5.59$, $s_2 = 5.45$ so $2.576\sqrt{\frac{(5.59)^2}{n} + \frac{(5.45)^2}{n}} = .5$ and thus $n = 1618$ students from each group should be sampled.

c. If it is required that all four sample sizes must be equal, use $n = 2998$ (from part a) to assure an interval width of 1 unit.

8.80 The 95% CI, based on a t–distribution with $21 - 1 = 20$ degrees of freedom, is

$$26.6 \pm 2.086\left(7.4/\sqrt{21}\right) = 26.6 \pm 3.37 \text{ or } (23.23, 29.97).$$

8.81 The sample statistics are $\bar{y} = 60.8$, $s = 7.97$. So, the 95% CI is

$$60.8 \pm 2.262\left(7.97/\sqrt{10}\right) = 60.8 \pm 5.70 \text{ or } (55.1, 66.5).$$

8.82 **a.** The 90% CI for the mean verbal SAT score for urban high school seniors is

$$505 \pm 1.729\left(57/\sqrt{20}\right) = 505 \pm 22.04 \text{ or } (482.96, 527.04).$$

b. Since the interval includes the score 508, it is a plausible value for the mean.

c. The 90% CI for the mean math SAT score for urban high school seniors is

$$495 \pm 1.729\left(69/\sqrt{20}\right) = 495 \pm 26.68 \text{ or } (468.32, 521.68).$$

The interval does include the score 520, so the interval supports the stated true mean value.

8.83 **a.** Using the sample–sample CI for $\mu_1 - \mu_2$, using an assumption of normality, we calculate the pooled sample variance

$$s_p^2 = \tfrac{9(3.92)^2+9(3.98)^2}{18} = 15.6034$$

Thus, the 95% CI for the difference in mean compartment pressures is

$$14.5–11.1 \pm 2.101\sqrt{15.6034\left(\tfrac{1}{10}+\tfrac{1}{10}\right)} = 3.4 \pm 3.7 \text{ or } (–.3, 7.1).$$

b. Similar to part a, the pooled sample variance for runners and cyclists who exercise at 80% maximal oxygen consumption is given by

$$s_p^2 = \tfrac{9(3.49)^2+9(4.95)^2}{18} = 18.3413.$$

The 90% CI for the difference in mean compartment pressures here is

$$12.2–11.5 \pm 1.734\sqrt{18.3413\left(\tfrac{1}{10}+\tfrac{1}{10}\right)} = .7 \pm 3.32 \text{ or } (–2.62, 4.02).$$

c. Since both intervals contain 0, we cannot conclude that the means in either case are different from one another.

8.84 The sample statistics are $\bar{y} = 3.781$, $s = .0327$. So, the 95% CI, with 9 degrees of freedom and $t_{.025} = 2.262$, is

$$3.781 \pm 2.262\left(.0327/\sqrt{10}\right) = 3.781 \pm .129 \text{ or } (3.652, 3.910).$$

8.85 The pooled sample variance is $s_p^2 = \tfrac{15(6)^2+19(8)^2}{34} = 51.647$. Then the 95% CI for $\mu_1 - \mu_2$ is

$$11-12 \pm 1.96\sqrt{51.647\left(\tfrac{1}{16}+\tfrac{1}{20}\right)} = –1 \pm 4.72 \text{ or } (–5.72, 3.72)$$

(here, we approximate $t_{.025}$ with $z_{.025} = 1.96$).

8.86 **a.** The sample statistics are, with $n = 14$, $\bar{y} = 0.896$, $s = .400$. The 95% CI for μ = mean price of light tuna in water, with 13 degrees of freedom and $t_{.025} = 2.16$ is

$$.896 \pm 2.16\left(.4/\sqrt{14}\right) = .896 \pm .231 \text{ or } (.665, 1.127).$$

b. The sample statistics are, with $n = 11$, $\bar{y} = 1.147$, $s = .679$. The 95% CI for μ = mean price of light tuna in oil, with 10 degrees of freedom and $t_{.025} = 2.228$ is

$$1.147 \pm 2.228\left(.679/\sqrt{11}\right) = 1.147 \pm .456 \text{ or } (.691, 1.603).$$

This CI has a larger width because: s is larger, n is smaller, $t_{\alpha/2}$ is bigger.

8.87 **a.** Following Ex. 8.86, the pooled sample variance is $s_p^2 = \tfrac{13(.4)^2+10(.679)^2}{23} = .291$. Then the 90% CI for $\mu_1 - \mu_2$, with 23 degrees of freedom and $t_{.05} = 1.714$ is

$$(.896 - 1.147) \pm 1.714\sqrt{.291\left(\tfrac{1}{14} + \tfrac{1}{11}\right)} = -.251 \pm .373 \text{ or } (-.624, .122).$$

b. Based on the above interval, there is not compelling evidence that the mean prices are different since 0 is contained inside the interval.

8.88 The sample statistics are, with $n = 12$, $\bar{y} = 9$, $s = 6.4$. The 90% CI for μ = mean LC50 for DDT is, with 11 degrees of freedom and $t_{.05} = 1.796$,

$$9 \pm 1.796\left(6.4/\sqrt{12}\right) = 9 \pm 3.32 \text{ or } (5.68, 12.32).$$

8.89 **a.** For the three LC50 measurements of Diazinon, $\bar{y} = 3.57$, $s = 3.67$. The 90% CI for the true mean is (2.62, 9.76).

b. The pooled sample variance is $s_p^2 = \frac{11(6.4)^2 + 2(3.57)^2}{13} = 36.6$. Then the 90% CI for the difference in mean LC50 chemicals, with 15 degrees of freedom and $t_{.025} = 1.771$, is

$$(9 - 3.57) \pm 1.771\sqrt{36.6\left(\tfrac{1}{12} + \tfrac{1}{3}\right)} = 5.43 \pm 6.92 \text{ or } (-1.49, 12.35).$$

We assumed that the sample measurements were independently drawn from normal populations with $\sigma_1 = \sigma_2$.

8.90 **a.** For the 95% CI for the difference in mean verbal scores, the pooled sample variance is $s_p^2 = \frac{14(42)^2 + 14(45)^2}{28} = 1894.5$ and thus

$$446 - 534 \pm 2.048\sqrt{1894.5\left(\tfrac{2}{15}\right)} = -88 \pm 32.55 \text{ or } (-120.55, -55.45).$$

b. For the 95% CI for the difference in mean math scores, the pooled sample variance is $s_p^2 = \frac{14(57)^2 + 14(52)^2}{28} = 2976.5$ and thus

$$548 - 517 \pm 2.048\sqrt{2976.5\left(\tfrac{2}{15}\right)} = 31 \pm 40.80 \text{ or } (-9.80, 71.80).$$

c. At the 95% confidence level, there appears to be a difference in the two mean verbal SAT scores achieved by the two groups. However, a difference is not seen in the math SAT scores.

d. We assumed that the sample measurements were independently drawn from normal populations with $\sigma_1 = \sigma_2$.

8.91 Sample statistics are:

Season	sample mean	sample variance	sample size
spring	15.62	98.06	5
summer	72.28	582.26	4

The pooled sample variance is $s_p^2 = \frac{4(98.06) + 3(582.26)}{7} = 305.57$ and thus the 95% CI is

$$15.62 - 72.28 \pm 2.365\sqrt{305.57\left(\tfrac{1}{5}+\tfrac{1}{4}\right)} = -56.66 \pm 27.73 \text{ or } (-84.39, -28.93).$$

It is assumed that the two random samples were independently drawn from normal populations with equal variances.

8.92 Using the summary statistics, the pooled sample variance is $s_p^2 = \frac{3(.001)+4(.002)}{7} = .0016$ and so the 95% CI is given by

$$.22 - .17 \pm 2.365\sqrt{.0016\left(\tfrac{1}{4}+\tfrac{1}{5}\right)} = .05 \pm .063 \text{ or } (-.013, .113).$$

8.93 **a.** Since the two random samples are assumed to be independent and normally distributed, the quantity $2\overline{X}+\overline{Y}$ is normally distributed with mean $2\mu_1 + \mu_2$ and variance $\left(\tfrac{4}{n}+\tfrac{3}{m}\right)\sigma^2$. Thus, is σ^2 is known, then $2\overline{X}+\overline{Y} \pm 1.96\sigma\sqrt{\tfrac{4}{n}+\tfrac{3}{m}}$ is a 95% CI for $2\mu_1 + \mu_2$.

b. Recall that $(1/\sigma^2)\sum_{i=1}^{n}(X_i - \overline{X})^2$ has a chi–square distribution with $n - 1$ degrees of freedom. Thus, $[1/(3\sigma^2)]\sum_{i=1}^{m}(Y_i - \overline{Y})^2$ is chi–square with $m - 1$ degrees of freedom and the sum of these is chi–square with $n + m - 2$ degrees of freedom. Then, by using Definition 7.2, the quantity

$$T = \frac{2\overline{X}+\overline{Y}-(2\mu_1+\mu_2)}{\hat{\sigma}\sqrt{\tfrac{4}{n}+\tfrac{3}{m}}}, \text{ where}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(X_i-\overline{X})^2 + \tfrac{1}{3}\sum_{i=1}^{m}(Y_i-\overline{Y})^2}{n+m-2}.$$

Then, the 95% CI is given by $2\overline{X}+\overline{Y} \pm t_{.025}\hat{\sigma}\sqrt{\tfrac{4}{n}+\tfrac{3}{m}}$.

8.94 The pivotal quantity is $T = \dfrac{\overline{Y}_1 - \overline{Y}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\tfrac{1}{n_1}+\tfrac{1}{n_2}}}$, which has a t–distribution w/ $n_1 + n_2 - 2$ degrees of freedom. By selecting t_α from this distribution, we have that $P(T < t_\alpha) = 1 - \alpha$. Using the same approach to derive the confidence interval, it is found that

$$\overline{Y}_1 - \overline{Y}_2 \pm t_\alpha S_p\sqrt{\tfrac{1}{n_1}+\tfrac{1}{n_2}}$$

is a $100(1 - \alpha)$% upper confidence bound for $\mu_1 - \mu_2$.

8.95 From the sample data, $n = 6$ and $s^2 = .503$. Then, $\chi^2_{.95} = 1.145476$ and $\chi^2_{.05} = 11.0705$ with 5 degrees of freedom. The 90% CI for σ^2 is $\left(\frac{5(.503)}{11.0705}, \frac{5(.503)}{1.145476}\right)$ or (.227, 2.196). We are 90% confident that σ^2 lies in this interval.

8.96 From the sample data, $n = 10$ and $s^2 = 63.5$. Then, $\chi^2_{.95} = 3.3251$ and $\chi^2_{.05} = 16.9190$ with 9 degrees of freedom. The 90% CI for σ^2 is $\left(\frac{571.6}{16.9190}, \frac{571.6}{3.3251}\right)$ or (33.79, 171.90).

8.97 **a.** Note that $1-\alpha = P\left(\frac{(n-1)S^2}{\sigma^2} > \chi^2_{1-\alpha}\right) = P\left(\frac{(n-1)S^2}{\chi^2_{1-\alpha}} > \sigma^2\right)$. Then, $\frac{(n-1)S^2}{\chi^2_{1-\alpha}}$ is a 100(1–α)% upper confidence bound for σ^2.

b. Similar to part (a), it can be shown that $\frac{(n-1)S^2}{\chi^2_{\alpha}}$ is a 100(1–α)% lower confidence bound for σ^2.

8.98 The confidence interval for σ^2 is $\left(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{\alpha/2}}\right)$, so since $S^2 > 0$, the confidence interval for σ is simply $\left(\sqrt{\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}}, \sqrt{\frac{(n-1)S^2}{\chi^2_{\alpha/2}}}\right)$.

8.99 Following Ex. 8.97 and 8.98:

a. 100(1 – α)% upper confidence bound for σ: $\sqrt{\frac{(n-1)S^2}{\chi^2_{1-\alpha}}}$.

b. 100(1 – α)% lower confidence bound for σ: $\sqrt{\frac{(n-1)S^2}{\chi^2_{\alpha}}}$.

8.100 With $n = 20$, the sample variance $s^2 = 34854.4$. From Ex. 8.99, a 99% upper confidence bound for the standard deviation σ is, with $\chi^2_{.99} = 7.6327$,

$$\sqrt{\frac{19(34854.4)}{7.6327}} = 294.55.$$

Since this is an upper bound, it is possible that the true population standard deviation is less than 150 hours.

8.101 With $n = 6$, the sample variance $s^2 = .0286$. Then, $\chi^2_{.95} = 1.145476$ and $\chi^2_{.05} = 11.0705$ with 5 degrees of freedom and the 90% CI for σ^2 is

$$\left(\frac{5(.0286)}{11.0705}, \frac{5(.0286)}{1.145476}\right) = (.013\ .125).$$

8.102 With $n = 5$, the sample variance $s^2 = 144.5$. Then, $\chi^2_{.995} = .20699$ and $\chi^2_{.005} = 14.8602$ with 4 degrees of freedom and the 99% CI for σ^2 is

$$\left(\frac{4(144.5)}{14.8602}, \frac{4(144.5)}{.20699}\right) = (38.90, 2792.41).$$

8.103 With $n = 4$, the sample variance $s^2 = 3.67$. Then, $\chi^2_{.95} = .351846$ and $\chi^2_{.05} = 7.81473$ with 3 degrees of freedom and the 99% CI for σ^2 is

$$\left(\frac{3(3.67)}{7.81473}, \frac{3(3.67)}{.351846}\right) = (1.4, 31.3).$$

An assumption of independent measurements and normality was made. Since the interval implies that the *standard deviation* could be larger than 5 units, it is possible that the instrument could be off by more than two units.

8.104 The only correct interpretation is choice d.

8.105 The difference of the endpoints 7.37 – 5.37 = 2.00 is equal to $2z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}} = 2z_{\alpha/2}\sqrt{\frac{6}{25}}$. Thus, $z_{\alpha/2} \approx 2.04$ so that $\alpha/2 = .0207$ and the confidence coefficient is $1 - 2(.0207) = .9586$.

8.106 **a.** Define: p_1 = proportion of survivors in low water group for male parents
p_2 = proportion of survivors in low nutrient group for male parents

Then, the sample estimates are $\hat{p}_1 = 522/578 = .903$ and $\hat{p}_2 = 510/568 = .898$. The 99% CI for the difference $p_1 - p_2$ is

$$.903 - .898 \pm 2.576\sqrt{\frac{.903(.097)}{578} + \frac{.898(.102)}{568}} = .005 \pm .0456 \text{ or } (-.0406, .0506).$$

b. Define: p_1 = proportion of male survivors in low water group
p_2 = proportion of female survivors in low water group

Then, the sample estimates are $\hat{p}_1 = 522/578 = .903$ and $\hat{p}_2 = 466/510 = .914$. The 99% CI for the difference $p_1 - p_2$ is

$$.903 - .914 \pm 2.576\sqrt{\frac{.903(.097)}{578} + \frac{.914(.086)}{510}} = -.011 \pm .045 \text{ or } (-.056, .034).$$

8.107 With $B = .03$ and $\alpha = .05$, we use the sample estimates of the proportions to solve

$$1.96\sqrt{\frac{.903(.097)}{n} + \frac{.898(.102)}{n}} = .03.$$

The solution is $n = 764.8$, therefore 765 seeds should be used in each environment.

8.108 If it is assumed that p = kill rate = .6, then this can be used in the sample size formula with $B = .02$ to obtain (since a confidence coefficient was not specified, we are using a multiple of 2 for the error bound)

$$.02 = 2\sqrt{\frac{.6(.4)}{n}}.$$

So, $n = 2400$.

8.109 **a.** The sample proportion of unemployed workers is 25/400 = .0625, and a two–standard–error bound is given by $2\sqrt{\frac{.0625(.9375)}{400}} = .0242$.

b. Using the same estimate of p, the true proportion of unemployed workers, gives the relation $2\sqrt{\frac{.0625(.9375)}{n}} = .02$. This is solved by $n = 585.94$, so 586 people should be sampled.

8.110 For an error bound of \$50 and assuming that the population standard deviation $\sigma = 400$, the equation to be solved is

$$1.96\frac{400}{\sqrt{n}} = 50.$$

This is solved by $n = 245.96$, so 246 textile workers should be sampled.

8.111 Assuming that the true proportion $p = .5$, a confidence coefficient of .95 and desired error of estimation $B = .005$ gives the relation

$$1.96\sqrt{\frac{.5(.5)}{n}} = .005.$$

The solution is $n = 38{,}416$.

8.112 The goal is to estimate the difference of

p_1 = proportion of all fraternity men favoring the proposition
p_2 = proportion of all non–fraternity men favoring the proposition

A point estimate of $p_1 - p_2$ is the difference of the sample proportions:

$$300/500 - 64/100 = .6 - .64 = -.04.$$

A two–standard–error bound is

$$2\sqrt{\frac{.6(.4)}{500} + \frac{.64(.36)}{100}} = .106.$$

8.113 Following Ex. 112, assuming equal sample sizes and population proportions, the equation that must be solved is

$$2\sqrt{\frac{.6(.4)}{n} + \frac{.6(.4)}{n}} = .05.$$

Here, $n = 768$.

8.114 The sample statistics are $\bar{y} = 795$ and $s = 8.34$ with $n = 5$. The 90% CI for the mean daily yield is

$$795 \pm 2.132\left(8.34/\sqrt{5}\right) = 795 \pm 7.95 \text{ or } (787.05, 802.85).$$

It was necessary to assume that the process yields follow a normal distribution and that the measurements represent a random sample.

8.115 Following Ex. 8.114 w/ $5 - 1 = 4$ degrees of freedom, $\chi^2_{.95} = .710721$ and $\chi^2_{.05} = 9.48773$. The 90% CI for σ^2 is (note that $4s^2 = 278$)

$$\left(\tfrac{278}{9.48773}, \tfrac{278}{.710721}\right) \text{ or } (29.30, 391.15).$$

8.116 The 99% CI for μ is given by, with 15 degrees of freedom and $t_{.005} = 2.947$, is

$$79.47 \pm 2.947\left(25.25/\sqrt{16}\right) = 79.47 \pm 18.60 \text{ or } (60.87, 98.07).$$

We are 99% confident that the true mean long–term word memory score is contained in the interval.

8.117 The 90% CI for the mean annual main stem growth is given by

$$11.3 \pm 1.746\left(3.4/\sqrt{17}\right) = 11.3 \pm 1.44 \text{ or } (9.86, 12.74).$$

8.118 The sample statistics are $\bar{y} = 3.68$ and $s = 1.905$ with $n = 6$. The 90% CI for the mean daily yield is

$$3.68 \pm 2.015\left(1.905/\sqrt{6}\right) = 3.68 \pm 1.57 \text{ or } (2.11, 5.25).$$

8.119 Since both sample sizes are large, we can use the large sample CI for the difference of population means:

$$75 - 72 \pm 1.96\sqrt{\tfrac{10^2}{50} + \tfrac{8^2}{45}} = 3 \pm 3.63 \text{ or } (–.63, 6.63).$$

8.120 Here, we will assume that the two samples of test scores represent random samples from normal distributions with $\sigma_1 = \sigma_2$. The pooled sample variance is $s_p^2 = \frac{10(52)+13(71)}{23} = 62.74$. The 95% CI for $\mu_1 - \mu_2$ is given by

$$64 - 69 \pm 2.069\sqrt{62.74\left(\tfrac{1}{11} + \tfrac{1}{14}\right)} = –5 \pm 6.60 \text{ or } (–11.60, 1.60).$$

8.121 Assume the samples of reaction times represent random sample from normal populations with $\sigma_1 = \sigma_2$. The sample statistics are: $\bar{y}_1 = 1.875$, $s_1^2 = .696$, $\bar{y}_2 = 2.625$, $s_2^2 = .839$. The pooled sample variance is $s_p^2 = \frac{7(.696)+7(.839)}{14} = .7675$ and the 90% CI for $\mu_1 - \mu_2$ is

$$1.875 - 2.625 \pm 1.761\sqrt{.7675\left(\tfrac{2}{8}\right)} = –.75 \pm .77 \text{ or } (–1.52, .02).$$

8.122 A 90% CI for μ = mean time between billing and payment receipt is, with $z_{.05} = 1.645$ (here we can use the large sample interval formula),

$$39.1 \pm 1.645\left(17.3/\sqrt{100}\right) = 39.1 \pm 2.846 \text{ or } (36.25, 41.95).$$

We are 90% confident that the true mean billing time is contained in the interval.

8.123 The sample proportion is 1914/2300 = .832. A 95% CI for p = proportion of all viewers who misunderstand is

$$.832 \pm 1.96\sqrt{\tfrac{.832(.168)}{2300}} = .832 \pm .015 \text{ or } (.817, .847).$$

8.124 The sample proportion is 278/415 = .67. A 95% CI for p = proportion of all corporate executives who consider cash flow the most important measure of a company's financial health is

$$.67 \pm 1.96\sqrt{\tfrac{.67(.33)}{415}} = .67 \pm .045 \text{ or } (.625, .715).$$

8.125 a. From Definition 7.3, the following quantity has an F–distribution with $n_1 - 1$ numerator and $n_2 - 1$ denominator degrees of freedom:

$$F = \frac{\frac{(n_1-1)S_1^2}{\sigma_1^2}\Big/(n_1 - 1)}{\frac{(n_2-1)S_2^2}{\sigma_2^2}\Big/(n_2 - 1)} = \frac{S_1^2}{S_2^2} \times \frac{\sigma_2^2}{\sigma_1^2}.$$

b. By choosing quantiles from the F–distribution with $n_1 - 1$ numerator and $n_2 - 1$ denominator degrees of freedom, we have

$$P(F_{1-\alpha/2} < F < F_{\alpha/2}) = 1 - \alpha.$$

Using the above random variable gives

$$P(F_{1-\alpha/2} < \frac{S_1^2}{S_2^2} \times \frac{\sigma_2^2}{\sigma_1^2} < F_{\alpha/2}) = P(\frac{S_2^2}{S_1^2}F_{1-\alpha/2} < \frac{\sigma_2^2}{\sigma_1^2} < \frac{S_2^2}{S_1^2}F_{\alpha/2}) = 1 - \alpha.$$

Thus,

$$\left(\frac{S_2^2}{S_1^2} F_{1-\alpha/2}, \frac{S_2^2}{S_1^2} F_{\alpha/2} \right)$$

is a 100(1 – α)% CI for σ_2^2 / σ_1^2.

An alternative expression is given by the following. Let $F_{\nu_2,\alpha}^{\nu_1}$ denote the upper–α critical value from the *F*–distribution with ν_1 numerator and ν_2 denominator degrees of freedom. Because of the relationship (see Ex. 7.29)

$$F_{\nu_2,\alpha}^{\nu_1} = \frac{1}{F_{\nu_1,\alpha}^{\nu_2}},$$

a 100(1 – α)% CI for σ_2^2 / σ_1^2 is also given by

$$\left(\frac{1}{F_{\nu_1,\alpha}^{\nu_2}} \frac{S_2^2}{S_1^2}, F_{\nu_2,\alpha}^{\nu_1} \frac{S_2^2}{S_1^2} \right).$$

8.126 Using the CI derived in Ex. 8.126, we have that $F_{9,.025}^{9} = \frac{1}{F_{9,.025}^{9}} = 4.03$. Thus, the CI for the ratio of the true population variances is $\left(\frac{1}{4.03} \cdot \frac{.094}{.273}, \frac{4.03(.094)}{.273} \right)$ = (.085, 1.39).

8.127 It is easy to show (e.g. using the mgf approach) that $\bar{Y}$ has a gamma distribution with shape parameter $100c_0$ and scale parameter (.01)β. In addition the statistic $U = \bar{Y} / \beta$ is a pivotal quantity since the distribution is free of β: the distribution of *U* is gamma with shape parameter $100c_0$ and scale parameter (.01). Now, $E(U) = c_0$ and $V(U) = (.01)c_0$ and by the Central Limit Theorem,

$$\frac{U - c_0}{.1\sqrt{c_0}} = \frac{\bar{Y}/\beta - c_0}{.1\sqrt{c_0}}$$

has an approximate standard normal distribution. Thus,

$$P\left(-z_{\alpha/2} < \frac{\bar{Y}/\beta - c_0}{.1\sqrt{c_0}} < z_{\alpha/2} \right) \approx 1 - \alpha.$$

Isolating the parameter β in the above inequality yields the desired result.

8.128 **a.** Following the notation of Section 8.8 and the assumptions given in the problem, we know that $\bar{Y}_1 - \bar{Y}_2$ is a normal variable with mean $\mu_1 - \mu_2$ and variance $\frac{\sigma_1^2}{n_1} + \frac{k\sigma_1^2}{n_2}$. Thus, the standardized variable Z^* as defined indeed has a standard normal distribution.

b. The quantities $U_1 = \frac{(n_1 - 1)S_1^2}{\sigma_1^2}$ and $U_2 = \frac{(n_2 - 1)S_2^2}{k\sigma_1^2}$ have independent chi–square distributions with $n_1 - 1$ and $n_2 - 1$ degrees of freedom (respectively). So, $W^* = U_1 + U_2$ has a chi–square distribution with $n_1 + n_2 - 2$ degrees of freedom.

c. By Definition 7.2, the quantity $T^* = \dfrac{Z^*}{\sqrt{W^*/(n_1+n_2-2)}}$ follows a t–distribution with $n_1 + n_2 - 2$ degrees of freedom.

d. A 100(1 – α)% CI for $\mu_1 - \mu_2$ is given by $\overline{Y}_1 - \overline{Y}_2 \pm t_{\alpha/2} S_p^* \sqrt{\frac{1}{n_1} + \frac{k}{n_2}}$, where $t_{\alpha/2}$ is the upper–α/2 critical value from the t–distribution with $n_1 + n_2 - 2$ degrees of freedom and S_p^* is defined in part (c).

e. If $k = 1$, it is equivalent to the result for $\sigma_1 = \sigma_2$.

8.129 Recall that $V(S^2) = \frac{2\sigma^4}{n-1}$.

a. $V(S'^2) = V(\frac{n-1}{n} S^2) = \frac{2(n-1)\sigma^4}{n^2}$.

b. The result follows from $V(S'^2) = V(\frac{n-1}{n} S^2) = \left(\frac{n-1}{n}\right)^2 V(S^2) < V(S^2)$ since $\frac{n-1}{n} < 1$.

8.130 Since S^2 is unbiased,

$\text{MSE}(S^2) = V(S^2) = \frac{2\sigma^4}{n-1}$. Similarly,

$\text{MSE}(S'^2) = V(S'^2) + [B(S'^2)]^2 = \frac{2(n-1)\sigma^4}{n^2} + \left(\frac{n-1}{n}\sigma^2 - \sigma^2\right)^2 = \frac{(2n-1)\sigma^4}{n^2}$.

By considering the ratio of these two MSEs, it can be seen that S'^2 has the smaller MSE and thus possibly a better estimator.

8.131 Define the estimator $\hat{\sigma}^2 = c\sum_{i=1}^{n}(Y_i - \overline{Y})^2$. Therefore, $E(\hat{\sigma}^2) = c(n-1)\sigma^2$ and $V(\hat{\sigma}^2) = 2c^2(n-1)\sigma^4$ so that

$$\text{MSE}(\hat{\sigma}^2) = 2c^2(n-1)\sigma^4 + [c(n-1)\sigma^2 - \sigma^2]^2.$$

Minimizing this quantity with respect to c, we find that the smallest MSE occurs when $c = \frac{1}{n+1}$.

8.132 **a.** The distribution function for $Y_{(n)}$ is given by

$$F_{Y_{(n)}}(y) = P(Y_{(n)} < y) = [F(y)]^n = \left(\frac{y}{\theta}\right)^{cn}, \; 0 \le y \le \theta.$$

b. The distribution of $U = Y_{(n)}/\theta$ is

$$F_U(u) = P(U \le u) = P(Y_{(n)} \le \theta u) = u^{nc}, \; 0 \le u \le 1.$$

Since this distribution is free of θ, $U = Y_{(n)}/\theta$ is a pivotal quantity. Also,

$$P\big(k < Y(n)/\theta \le 1\big) = P\big(k\theta < Y(n) \le \theta\big) = F_{Y_{(n)}}(\theta) - F_{Y_{(n)}}(k\theta) = 1 - k^{cn}.$$

c. i. Using the result from part b with $n = 5$ and $c = 2.4$,

$.95 = 1 - (k)^{12}$ so $k = .779$

ii. Solving the equations .975 = 1 – $(k_1)^{12}$ and .025 = 1 – $(k_2)^{12}$, we obtain k_1 = .73535 and k_2 = .99789. Thus,

$$P\left(.73535 < Y_{(5)}/\theta < .99789\right) = P\left(\frac{Y_{(5)}}{.99789} < \theta < \frac{Y_{(5)}}{.73535}\right) = .95.$$

So, $\left(\frac{Y_{(5)}}{.99789}, \frac{Y_{(5)}}{.73535}\right)$ is a 95% CI for θ.

8.133 We know that $E(S_i^2) = \sigma^2$ and $V(S_i^2) = \frac{2\sigma^2}{n_i - 1}$ for i = 1, 2.

a. $E(S_p^2) = \dfrac{(n_1-1)E(S_1^2) + (n_2-1)E(S_2^2)}{n_1+n_2-2} = \sigma^2$

b. $V(S_p^2) = \dfrac{(n_1-1)^2 V(S_1^2) + (n_2-1)^2 V(S_2^2)}{(n_1+n_2-2)^2} = \dfrac{2\sigma^4}{n_1+n_2-2}$.

8.134 The width of the small sample CI is $2t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right)$, and from Ex. 8.16 it was derived that $E(S) = \frac{\sigma}{\sqrt{n-1}}\frac{\sqrt{2}\Gamma(n/2)}{\Gamma[(n-1)/2]}$. Thus,

$$E\left(2t_{\alpha/2}\frac{S}{\sqrt{n}}\right) = 2^{3/2}t_{\alpha/2}\left(\frac{\sigma}{\sqrt{n(n-1)}}\right)\left(\frac{\Gamma(n/2)}{\Gamma[(n-1)/2]}\right).$$

8.135 The midpoint of the CI is given by $M = \frac{1}{2}\left(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}} + \frac{(n-1)S^2}{\chi^2_{\alpha/2}}\right)$. Therefore, since $E(S^2) = \sigma^2$, we have

$$E(M) = \frac{1}{2}\left(\frac{(n-1)\sigma^2}{\chi^2_{1-\alpha/2}} + \frac{(n-1)\sigma^2}{\chi^2_{\alpha/2}}\right) = \frac{(n-1)\sigma^2}{2}\left(\frac{1}{\chi^2_{1-\alpha/2}} + \frac{1}{\chi^2_{\alpha/2}}\right) \neq \sigma^2.$$

8.136 Consider the quantity $Y_p - \bar{Y}$. Since $Y_1, Y_2, \ldots, Y_n, Y_p$ are independent and identically distributed, we have that

$$E(Y_p - \bar{Y}) = \mu - \mu = 0$$
$$V(Y_p - \bar{Y}) = \sigma^2 + \sigma^2/n = \sigma^2\left(\tfrac{n+1}{n}\right).$$

Therefore, $Z = \dfrac{Y_p - \bar{Y}}{\sigma\sqrt{\frac{n+1}{n}}}$ has a standard normal distribution. So, by Definition 7.2,

$$\frac{\dfrac{Y_p - \bar{Y}}{\sigma\sqrt{\frac{n+1}{n}}}}{\sqrt{\dfrac{(n-1)S^2}{\sigma^2(n-1)}}} = \frac{Y_p - \bar{Y}}{S\sqrt{\frac{n+1}{n}}}$$

has a t–distribution with n – 1 degrees of freedom. Thus, by using the same techniques as used in Section 8.8, the *prediction interval* is

$$\bar{Y} \pm t_{\alpha/2}S\sqrt{\tfrac{n+1}{n}},$$

where $t_{\alpha/2}$ is the upper–$\alpha/2$ critical value from the t–distribution with $n - 1$ degrees of freedom.