CHAPTER 10

Section 10.1

1.

a. $E(\overline{X}-\overline{Y}) = E(\overline{X}) - E(\overline{Y}) = 4.1 - 4.5 = -.4$, irrespective of sample sizes.

b. $V(\overline{X}-\overline{Y}) = V(\overline{X}) + V(\overline{Y}) = \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n} = \frac{(1.8)^2}{100} + \frac{(2.0)^2}{100} = .0724$, and the SD of $\overline{X}-\overline{Y}$ is $\overline{X}-\overline{Y} = \sqrt{.0724} = .2691$.

c. A normal curve with mean and SD as given in **a** and **b** (because $m = n = 100$, the CLT implies that both $\overline{X}$ and $\overline{Y}$ have approximately normal distributions, so $\overline{X}-\overline{Y}$ does also). The shape is not necessarily that of a normal curve when $m = n = 10$, because the CLT cannot be invoked. So if the two lifetime population distributions are not normal, the distribution of $\overline{X}-\overline{Y}$ will typically be quite complicated.

3.

a. The test statistic value is $z = \frac{(\overline{x}-\overline{y})-0}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}}$, and H_0 will be rejected at level .05 if $|z| \geq$ 1.96. We compute $z = \frac{(42,500-40,400)-0}{\sqrt{\frac{2200^2}{45}+\frac{1900^2}{45}}} = \frac{2100}{433.33} = 4.84 \geq 1.96$, so we reject H_0 and conclude that the true average tread lives for these two tire brands differ.

b. CI = $(\overline{x}-\overline{y}) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}$ = 2,100 ± 1.96(433.33) = (1251, 2949). As a practical matter, this is a fairly wide interval, suggesting $\mu_1 - \mu_2$ has not been estimated very precisely.

5.

a. H_a says that the average calorie output for sufferers is more than 1 cal/cm^2/min below that for non-sufferers. $\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}} = \sqrt{\frac{(.2)^2}{10}+\frac{(.4)^2}{10}} = .1414$, so $z = \frac{(.64-2.05)-(-1)}{.1414} = -2.90$. At level .01, H_0 is rejected if $z \leq -2.33$; since $-2.90 \leq -2.33$, reject H_0.

b. From **a,** P-value = $\Phi(-2.90)$ = .0019.

c. $\beta = 1 - \Phi\left(-2.33 - \frac{-1.2+1}{.1414}\right) = 1 - \Phi(-.92) = .8212$. Power = $1 - \beta$ = 1 – .8212 = .1788.

d. $m = n = \frac{.2(2.33+1.28)^2}{(-1.2-(-1))^2} = 65.15$, so use 66.

7.

a. Due to the relatively small sample sizes, we must assume here that the population elapsed time distributions are both normal. The Central Limit Theorem can't rescue us here.

b. To test H_0: $\mu_1 - \mu_2 = 0$ vs H_a: $\mu_1 - \mu_2 \neq 0$, we reject H_0 at the .01 level if $|z| \geq z_{.005} = 2.576$. The observed test statistic value is $z = \frac{(30.42 - 26.53) - 0}{\sqrt{8.5^2/15 + 8.5^2/19}} = \frac{3.89}{2.936} = 1.32$. Since $|1.32| < 2.576$, H_0 isn't rejected at the .01 level. The data do not provide convincing statistical evidence that the true average times differ.

9. $\sigma_1 = \sigma_2 = .2$, $\alpha = \beta = .05$, and the test is one-tailed, so

$n = \frac{\left(.2^2 + .2^2\right)\left(1.645 + 1.645\right)^2}{(.2-0)^2} = 21.65$. Use $n = 22$ hospitals of each type. We cannot make cause-and-effect conclusions here, since this is merely an observational study (nurse staffing problems were not forcibly introduced into randomly selected hospitals!). The general financial state of a hospital may impact both its nursing staff and its mortality rate.

11.

a. As either m or n increases, SD decreases, so $\frac{\mu_1 - \mu_2 - \Delta_0}{SD}$ increases (the numerator is positive), so $\left(z_\alpha - \frac{\mu_1 - \mu_2 - \Delta_0}{SD}\right)$ decreases, so $\beta = \Phi\left(z_\alpha - \frac{\mu_1 - \mu_2 - \Delta_0}{SD}\right)$ decreases.

b. As β decreases, z_β increases, and since z_β is in the numerator of n, n increases also.

Section 10.2

13.

a. $\nu = \frac{\left(\frac{5^2}{10} + \frac{6^2}{10}\right)^2}{\frac{\left(\frac{5^2}{10}\right)^2}{9} + \frac{\left(\frac{6^2}{10}\right)^2}{9}} = \frac{37.21}{.694 + 1.44} = 17.43 \approx 17.$

b. $\nu = \frac{\left(\frac{5^2}{10} + \frac{6^2}{15}\right)^2}{\frac{\left(\frac{5^2}{10}\right)^2}{9} + \frac{\left(\frac{6^2}{15}\right)^2}{14}} = \frac{24.01}{.694 + .411} = 21.7 \approx 21.$

c. $\nu = \frac{\left(\frac{2^2}{10} + \frac{6^2}{15}\right)^2}{\frac{\left(\frac{2^2}{10}\right)^2}{9} + \frac{\left(\frac{6^2}{15}\right)^2}{14}} = \frac{7.84}{.018 + .411} = 18.27 \approx 18.$

d. $\nu = \dfrac{\left(\frac{5^2}{12}+\frac{6^2}{24}\right)^2}{\dfrac{\left(\frac{5^2}{12}\right)^2}{11}+\dfrac{\left(\frac{6^2}{24}\right)^2}{23}} = \dfrac{12.84}{.395+.098} = 26.05 \approx 26.$

15. Let μ_1 and μ_2 denote the true mean years of education for all sons of foreign-born and native-born fathers in Germany, respectively. The goal is to test H_0: $\mu_1 - \mu_2 = 0$ vs H_a: $\mu_1 - \mu_2 < 0$; the latter is equivalent to the statement $\mu_2 > \mu_1$. With such large sample sizes, we may use a z approximation to the two-sample t test; in particular, H_0 will be rejected if $t \le -z_{.01} = -2.33$. The observed test statistic value is $t = \dfrac{(9.2-11.7)-0}{\sqrt{1.9^2/251+2.6^2/640}} = -15.83$ (a massive test statistic). In particular, since $-15.83 \le -2.33$, H_0 is resoundingly rejected at the .01 level. The data provide overwhelming evidence that the true average years of education for sons of native-born fathers in Germany exceeds that of sons with foreign-born fathers.

17. We will assume throughout this analysis that the relevant distributions are approximately normal.

a. Let μ_1 and μ_2 denote true mean head acceleration (g) with a helmet and with no helmet, respectively. The hypotheses of interest are H_0: $\mu_1 - \mu_2 = 0$ vs H_a: $\mu_1 - \mu_2 < 0$. Welch's df are roughly $\nu = 38$, and H_0 will be rejected if $t \le -t_{.05,38} = -1.686$. Here, $t = \dfrac{(43.1-75.4)-0}{\sqrt{4.5^2/24+7.2^2/24}} = -18.64 \le -1.686$, and we resoundingly reject the null hypothesis. The data provide overwhelming evidence that mean head acceleration is reduced with helmets.

b. Now let μ_1 and μ_2 denote true resultant neck force (N) with a helmet and with no helmet, respectively. The hypotheses of interest are H_0: $\mu_1 - \mu_2 = 0$ vs H_a: $\mu_1 - \mu_2 > 0$. Now Welch's df are $\nu = 44$, and H_0 will be rejected if $t \ge t_{.05,44} = 1.680$. Here, $t = \dfrac{(1331-945)-0}{\sqrt{93^2/24+77^2/24}} = 15.66 \ge 1.680$, and we resoundingly reject the null hypothesis. The data provide convincing evidence that the true mean resultant neck force is greater with helmets than without.

c. P(at least one type I error) $\le P$(type I error in (a)) + P(type I error in (b)) = .05 + .05 = .10. That is, the chance of committing at least one type I error is *at most* 10%.

19.

a. A 95% confidence interval for the fast food mean – non fast food mean = $\mu_2 - \mu_1$ is

$$\left(\bar{x}_2 - \bar{x}_1\right) \pm 1.96\sqrt{\frac{s_2^2}{n}+\frac{s_1^2}{m}} = \left(2637-2258\right) \pm 1.96\sqrt{\frac{1138^2}{413}+\frac{1519^2}{663}} = (219.6, 538.4)$$

. [The very large sample sizes imply that a z critical value is suitable here.]

b. We wish to test H_0: $\mu_2 - \mu_1 = 200$ vs H_a: $\mu_2 - \mu_1 > 200$. Given the large sample sizes, we will reject H_0 if $t \geq z_{.05} = 1.645$. Here, $t = \dfrac{(2637 - 2258) - 200}{\sqrt{\dfrac{1138^2}{413} + \dfrac{1519^2}{663}}} = \dfrac{179}{81.338} = 2.20 \geq 1.645$.

Equivalently, the one-tailed P-value is roughly $1 - \Phi(2.20) = .014$, which is less than .05. So reject the null hypothesis at the .05 level, and conclude that yes, there is strong evidence of a difference in means exceeding 200 calories per day.

21.

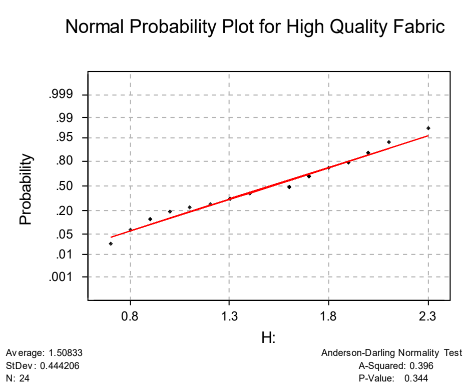
a. No, these distributions cannot be normal. In both samples, dollar values cannot be negative, but the sd exceeds the mean. So, both distributions must be positively skewed. (This also makes intuitive sense.) However, since we have such large samples, the sampling distributions of the two sample means are normal anyway; normally distributed populations are not vital to this analysis.

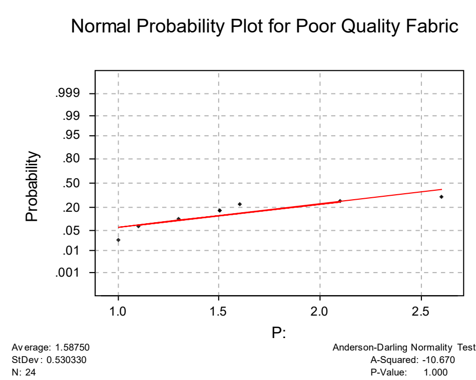
b. We build a 95% CI for this population difference: $(666 - 421) \pm \sqrt{\dfrac{1048^2}{75} + \dfrac{686^2}{209}} = \$245$ $\pm \$130 = (\$115, \$375)$. With 95% confidence, the mean account balance for students whose parents helped acquire a credit card is between \$115 and \$375 *higher* than the mean for students whose parents had no involvement whatsoever.

23. With sample 1 begin amateurs and sample 2 being professionals, we wish to test the hypotheses H_0: $\mu_1 = \mu_2$ versus H_a: $\mu_1 < \mu_2$. Calculating df as in the text gives $\nu = 42$, and the test statistic is $t = \dfrac{74.5 - 81.8}{\sqrt{6.29^2/24 + 8.64^2/24}} = -3.35$ The one-sided P-value is $P(T \leq -3.35) \approx$ .001, using the df = 40 column of the t table. So we reject H_0 and conclude that, on average, expert pianists hit the keys harder than amateur pianists.

25.

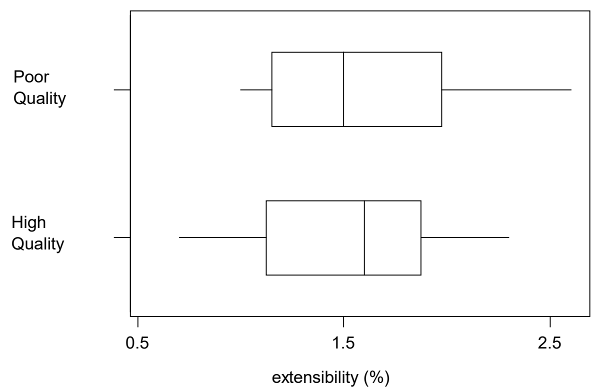
a. We see that both plots illustrate sufficient linearity. Therefore, it is plausible that both samples have been selected from normal population distributions.





b. The comparative boxplot does not suggest a difference between average extensibility for the two types of fabrics.

Comparative Box Plot for High Quality and Poor Quality Fabric



c. We test $H_0: \mu_1 - \mu_2 = 0$ vs. $H_a: \mu_1 - \mu_2 \neq 0$. With degrees of freedom

$\nu = \dfrac{(.0433265)^2}{.00017906} = 10.5$, which we round down to 10, and using significance level .05 (not specified in the problem), we reject H_0 if $|t| \geq t_{.025,10} = 2.228$. The test statistic is

$t = \dfrac{-.08}{\sqrt{(.0433265)}} = -.38$, which is not ≥ 2.228 in absolute value, so we cannot reject H_0.

There is insufficient evidence to claim that the true average extensibility differs for the two types of fabrics.

27. The null hypothesis is H_0: $\mu_1 = \mu_2$ and the alternative hypothesis is H_a: $\mu_1 < \mu_2$. Compute the test statistic $t = \dfrac{75.6 - 79.6}{\sqrt{5.9^2/40 + 7.6^2/40}} = -2.63$. The approximate degrees of freedom are

$\nu = \dfrac{(5.9^2/40 + 7.6^2/40)^2}{\dfrac{(5.9^2/40)^2}{39} + \dfrac{(7.6^2/40)^2}{39}} = 73.48$, which we round down to 73. The P-value for our lower tailed test is then .005, so at the .01 level we conclude that the true average range of motion for the pitchers is less than that for the position players. This claim could be false, in which case we have made a type I error.

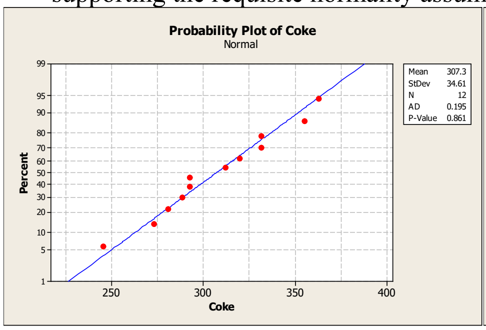
29. We will test the hypotheses: H_0: $\mu_1 - \mu_2 = 10$ v. H_a: $\mu_1 - \mu_2 > 10$. The test statistic is

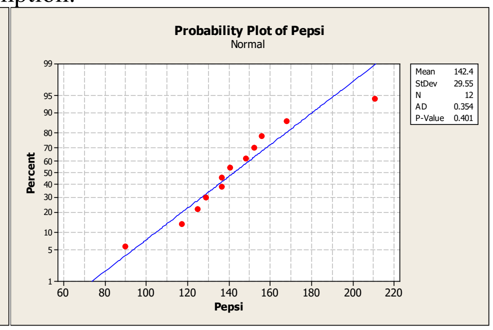
$t = \dfrac{(\bar{x} - \bar{y}) - 10}{\sqrt{\left(\frac{2.75^2}{10} + \frac{4.44^2}{5}\right)}} = \dfrac{4.5}{2.17} = 2.08$ with df $= \nu = \dfrac{\left(\frac{2.75^2}{10} + \frac{4.44^2}{5}\right)^2}{\dfrac{\left(\frac{2.75^2}{10}\right)^2}{9} + \dfrac{\left(\frac{4.44^2}{5}\right)^2}{4}} = \dfrac{22.08}{3.95} = 5.59 \searrow 5$, and the

P-value from the t table is approximately .045, which is < .10 so we reject H_0 and conclude that the true average lean angle for older females is more than 10 degrees smaller than that of younger females.

31.

a. Probability plots for the Coke and Pepsi data appear below. Both are fairly linear, supporting the requisite normality assumption.





b. The mean and sd for the Coke data are 307.28 and 34.61, while the mean and sd for the Pepsi data are 142.44 and 29.55. The estimated degrees of freedom are $\nu = 21$, and the t critical value is $t_{.005,21} = 2.831$. The resulting 99% CI for the difference in population means is (127.63, 202.03).

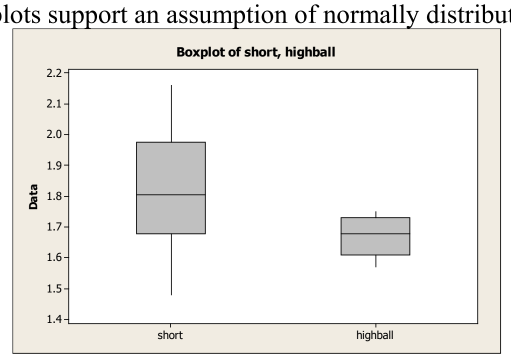
c. No. For a 99% lower confidence bound we use (difference of means) – $t_{.01,21}$(se) = 131.75.

d. We are 99% confident that the average foam volume from a 12 oz can of Coke is at least 131.75ml greater than the average foam volume from a 12 oz can of Pepsi.

33. Let μ_1 = the true average proportional stress limit for red oak and let μ_2 = the true average proportional stress limit for Douglas fir. We test $H_0: \mu_1 - \mu_2 = 1$ vs. $H_a: \mu_1 - \mu_2 > 1$. The test statistic is $t = \dfrac{(8.48 - 6.65) - 1}{\sqrt{\frac{.79^2}{14} + \frac{1.28^2}{10}}} = 1.818$. With degrees of freedom $\nu \approx 13.85 \to 13$, the P-value = $P(T > 1.8) = .048$. At $\alpha = .05$, there is sufficient evidence to claim that true average proportional stress limit for red oak exceeds that of Douglas fir by more than 1 MPa.

35.

a. It appears that bartenders pour slightly less rum into highball glasses, on average. But the most stark difference is variability: the amount poured into a slender, highball glass is much more consistent across bartenders than the amount poured into short, tumbler glasses. Both boxplots support an assumption of normally distributed populations.



b. As noted above, the two samples appear normal; probability plots confirm this. Software reports the following: $t = 1.88$ with estimated df = 8. The corresponding two-sided P-value from software is 0.097; hence, we fail to reject the null hypothesis at the standard $\alpha = 0.05$ level. We conclude that the true average amount of rum poured by experienced bartenders does not differ significantly from tumblers to highball glasses.

37. Let μ_1 and μ_2 be the average OCSD scores for the appropriate populations of males and females, respectively. We wish to test H_0: $\mu_1 = \mu_2$ versus H_a: $\mu_1 \neq \mu_2$. The samples are moderate in size, so, we use a two-sample t test. Software gives the following results: $t = 2.19$, estimated df = 81, P-value = .031. Hence, we reject the null hypothesis at the standard $\alpha = .05$. At this level, we conclude that the average OCSD scores are different for the populations of males and females with comorbid alcohol addiction and PTSD. If we use the stricter $\alpha = .01$ standard instead, we would fail to reject H_0, because .031 > .01.

39. As suggested in the hint, start with the facts $(m-1)S_1^2/\sigma^2 \sim \chi^2_{m-1}$ and $(n-1)S_2^2/\sigma^2 \sim \chi^2_{n-1}$. Since the X and Y samples are independent, so are their sample variances, which implies that the sum of the two terms above is also a chi-squared rv (sum of independent chi-squares is chi-squared), with df = $(m-1) + (n-1) = m + n - 2$. Put it all together:

$$\frac{(m+n-2)S_p^2}{\sigma^2} = \frac{(m-1)S_1^2}{\sigma^2} + \frac{(n-1)S_2^2}{\sigma^2} \sim \chi^2_{m-1} + \chi^2_{n-1} = \chi^2_{m+n-2}\,.$$

41.

a. Let $t = t_{\alpha/2,m+n-2}$. Then $1 - \alpha = P(-t < T < t)$, where T is the rv from the previous exercise. Solve the system of inequalities for $\mu_1 - \mu_2$:

$$-t < \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_p\sqrt{1/m+1/n}} < t \Leftrightarrow (\bar{X}-\bar{Y}) - t\cdot S_p\sqrt{1/m+1/n} < \mu_1-\mu_2 < (\bar{X}-\bar{Y}) + t\cdot S_p\sqrt{1/m+1/n}$$

Therefore, a "pooled" CI for $\mu_1 - \mu_2$ has endpoints $(\bar{x}-\bar{y}) \pm t_{\alpha/2,m+n-2}\cdot s_p\sqrt{1/m+1/n}$.

b. We have $m = 10$, $\bar{x} = 2903$, $s_1 = 277$, $n = 8$, $\bar{y} = 3108$, $s_2 = 206$. The t critical value is $t_{.025,10+8-2} = 2.120$, and the pooled variance is $s_p^2 = \frac{10-1}{10+8-2}(277)^2 + \frac{8-1}{10+8-2}(206)^2 =$ 61725.8, so $s_p = 248$. The resulting 95% CI for $\mu_1 - \mu_2$ is $(2903-3108) \pm 2.120\cdot 248\sqrt{1/10+1/8} = (-455, 45)$.

c. Without pooling, we need Welch's df, which here is $\nu = 15$. The traditional two-sample t CI for $\mu_1 - \mu_2$ is $(\bar{x}-\bar{y}) \pm t_{\alpha/2,\nu}\cdot\sqrt{s_1^2/m + s_2^2/n}$ = $(2903-3108) \pm 2.131\sqrt{277^2/10 + 206^2/8}$ = $(-448, 38)$. The two CI's are fairly close to each other.

Section 10.3

43. $\bar{d} = 13.33$, $s_D = 18.41$

1 Parameter of Interest: μ_D = true average difference

2 $H_0 : \mu_D = 0$ versus $H_a : \mu_D > 0$

3 A normal plot (not shown) is sufficiently straight to support normality for the population of differences.

4 $t = \dfrac{\bar{d} - \mu_D}{s_D / \sqrt{n}} = \dfrac{\bar{d} - 0}{s_D / \sqrt{n}}$

5 rejection region: $t \ge t_{.01,5} = 3.365$

6 $t = \dfrac{13.33 - 0}{8.41 / \sqrt{6}} = 3.88$

7 Reject H_0, and conclude that true average movement for the TightRope treatment is indeed less than that for the Fiber Mesh treatment.

45.

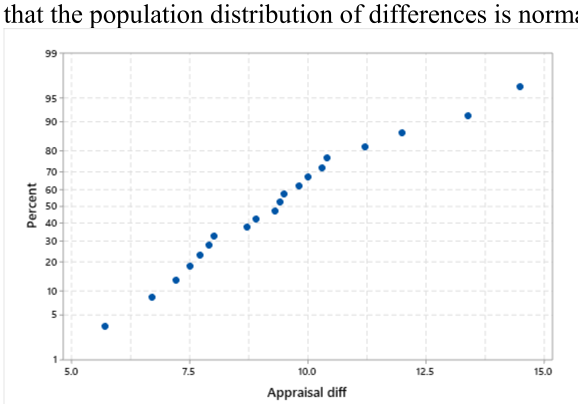
a. Let μ_D denote the population mean difference. From software, the mean and sd of the differences are .000246 and .000331, respectively. The t critical value is $t_{.025,12} = 2.179$, and the resulting 95% CI for μ_D is (.000046, .000446). Because 0 is not included in this interval, it does appear that the shovels differ with respect to true average energy expenditure, and that the difference is positive, so true energy expenditure with the conventional shovel is higher.

b. Compute $t = \dfrac{.000246 - 0}{.000331 / \sqrt{13}} = 2.68$ and the corresponding one-tailed P-value = .01.

Because this is less than .05 we reject the null hypothesis of equal population means, and conclude at the .05 level that true average energy expenditure using the conventional shovel exceeds that using the perforated shovel.

47.

a. The accompanying normal probability plot is quite linear, suggesting it is indeed plausible that the population distribution of differences is normal.



b. $\bar{d} + t_{.05,20-1}s_D/\sqrt{n} = 9.405 + (1.729)2.196/\sqrt{20}$ = 10.524. We can be 95% confident that the true mean difference in appraisal values is *at most* \$10,524.

c. To test H_0: $\mu_D = 10$ vs H_a: $\mu_D < 10$ at the .05 level, we will reject H_0 if $t \le -t_{.05,20-1} = -1.729$. The test statistic value is $t = \dfrac{\bar{d} - \mu_0}{s_D/\sqrt{n}} = \dfrac{9.405 - 10}{2.196/\sqrt{20}} = -1.21 > -1.729$, so we fail to reject H_0 here. This is consistent with part **b**: we concluded that $\mu_D < 10.524$, but that does not necessarily imply that $\mu_D < 10$. Thus, we fail to reject H_0: $\mu_D = 10$ in favor of H_a: $\mu_D < 10$ at .05 significance (aka 95% confidence).

49.

a. The two samples of 23 students are not matched or paired in any way. Rather, they may be regarded as two independent samples of students. Thus, the two-sample *t* procedures from Section 10.2 are appropriate here.

b. Let μ_1 and μ_2 denote the true mean payment offer in the 7 oz and 8 oz conditions, respectively. The goal is to test H_0: $\mu_1 - \mu_2 = 0$ vs H_a: $\mu_1 - \mu_2 > 0$. Welch's df are roughly ν = 43, and so H_0 will be rejected if $t \ge t_{.05,43} = 1.681$. Here, $t = \dfrac{(2.26 - 1.66) - 0}{\sqrt{.84^2/23 + .81^2/23}} = 2.47 \ge 1.681$, so we reject H_0. The data provide convincing statistical evidence that people will pay more, on average, for 7 oz of ice cream in a 5 oz cup than for 8 oz of ice cream in a 10 oz cup.

c. Now each of 23 students makes two offers, so the data are naturally paired by student. Hence, a paired *t* test is appropriate for this part of the analysis.

d. With $n = 23$, me must assume the population difference distribution is at least approximately normal. The summary statistics are $\bar{d} = \bar{x} - \bar{y} = 1.56 - 1.85 = -.29$ and $s_D = .32$. To test H_0: $\mu_D = 0$ vs H_a: $\mu_D < 0$ at the .05 level, we will reject H_0 if $t \le -t_{.05,23-1} = -1.717$. Here, $t = \dfrac{-.29 - 0}{.32/\sqrt{23}} = -4.34 \le -1.717$, so we reject H_0. At the 5% level, the data affirm the researchers' theory that students will offer to pay more, on average, for 8 oz of ice cream than for 7 oz.

e. In the first part of the study, students' only point of reference is the cup — most people can't distinguish 7 oz from 8 oz in the abstract. So, the overflowing ice cream cup looks like more ice cream. But when the two are side by side, even with different-sized cups, clearly one can see which ice cream volume is greater!

51.

a. The normal probability plot is very linear, except for the two lowest values. But these do not really weaken the plausibility of a normal population of differences.

b. For these differences, $n = 33$, $\bar{d} = 19.92$, $s_d = 17.61$. Thus, a 95% CI for μ_d is $\bar{d} \pm t_{.025,32}\left(\frac{s_d}{\sqrt{n}}\right) = 19.92 \pm (2.037)\left(\frac{17.61}{\sqrt{33}}\right)$ = (12.67, 25.16). We are 95% confident that the true mean difference in mumbling from first grade to third grade is between 12.67% and 25.16%.

53.

1 Parameter of interest: μ_D denotes the true average difference of spatial ability in brothers exposed to DES and brothers not exposed to DES.

2 $H_0 : \mu_D = 0$

3 $H_a : \mu_D > 0$

4 $t = \dfrac{\bar{d} - \mu_D}{s_D / \sqrt{n}} = \dfrac{\bar{d} - 0}{s_D / \sqrt{n}}$

5 RR: P-value < .05, df = 9

6 $t = \dfrac{(12.6 - 13.7) - 0}{0.5} = -2.2$, with corresponding P-value .028

7 Reject H_0. The data supports the idea that exposure to DES reduces spatial ability.

55. With $(x_1, y_1) = (6,5)$, $(x_2, y_2) = (15,14)$, $(x_3, y_3) = (1,0)$, and $(x_4, y_4) = (21,20)$, $\bar{d} = 1$ and $s_D = 0$ (the d_i's are 1, 1, 1, and 1), so the paired t statistic would be infinite. Meanwhile, $s_1 = s_2 = 8.96$ and $t = .16$ if we incorrectly apply the two-sample t procedure.

Section 10.4

57.

a. H_0 will be rejected if $|z| \geq 1.96$. With $\hat{p}_1 = \frac{63}{300} = .2100$, and $\hat{p}_2 = \frac{75}{180} = .4167$,

$\hat{p} = \dfrac{63 + 75}{300 + 180} = .2875$, $z = \dfrac{.2100 - .4167}{\sqrt{(.2875)(.7125)\left(\frac{1}{300} + \frac{1}{180}\right)}} = \dfrac{-.2067}{.0427} = -4.84$. Since $|-4.94| \geq 1.96$, H_0 is rejected.

b. $\bar{p} = .275$ and $\hat{p}_1 = .150$, so power =

$$1 - \left[\Phi\left(\frac{[(1.96)(.0421) + .2]}{.0432}\right) - \Phi\left(\frac{[-(1.96)(.0421) + .2]}{.0432}\right)\right] =$$

$$1 - [\Phi(6.54) - \Phi(2.72)] = .9967 .$$

59. Let α = .05. A 95% confidence interval is $\left(\hat{p}_1 - \hat{p}_2\right) \pm z_{\alpha/2}\sqrt{\left(\frac{\hat{p}_1\hat{q}_1}{m} + \frac{\hat{p}_2\hat{q}_2}{n}\right)}$

$$= \left(\tfrac{224}{395} - \tfrac{126}{266}\right) \pm 1.96\sqrt{\left(\frac{\left(\frac{224}{395}\right)\left(\frac{171}{395}\right)}{395} + \frac{\left(\frac{126}{266}\right)\left(\frac{140}{266}\right)}{266}\right)} = .0934 \pm .0774 = \left(.0160, .1708\right).$$

61.

a. With $\hat{p}_1$ = 322/1785 = .180 and $\hat{p}_2$ = 511/1186 = .431, a 99% CI for $p_1 - p_2$ is given by

$$(.180 - .431) \pm 2.576\sqrt{\frac{.180(1-.180)}{1785} + \frac{.431(1-.431)}{1186}} = (-.294, -.207).$$

b. Food images in British TV commercials are much more likely to include sugary and/or fatty foods than images in TV programs. In particular, the proportion of all commercials with sugary/fatty food images is between .207 and .294 higher than the proportion of all programs with sugary/fatty food images.

63. Let p_1 = the proportion of all Chinese brands in low-uncertainty business environments that use a "lucky" number of strokes, and let p_2 = the corresponding proportion for high-uncertainty brands. The researchers' hypotheses are H_0: $p_1 - p_2 = 0$ v H_a: $p_1 - p_2 < 0$. With $\hat{p}_1$ = 372/654 = .569, $\hat{p}_2$ = 343/548 = .626, and $\hat{p}$ = (372 + 343)/(654 + 548) = .595, the test statistic value is $z = \dfrac{(.569 - .626) - 0}{\sqrt{.595(1-.595)(1/654 + 1/548)}}$ = –2.01. The lower-tailed P-value is $P(Z \le -2.01) = \Phi(2.01) = .022 < \alpha = .05$; equivalently, $z = -2.01 \le -z_{.05} = -1.645$. Either way, H_0 is rejected at the .05 level, meaning the sample data support the researchers' theory.

65.

a. Let p_1 = the proportion of all students who would agree to be surveyed by Melissa and let p_2 = the proportion of all students who would agree to be surveyed by Kristine. The hypotheses of interest are H_0: $p_1 - p_2 = 0$ v H_a: $p_1 - p_2 \ne 0$. With $\hat{p}_1$ = 41/50 = .82, $\hat{p}_2$ = 27/50 = .54, and $\hat{p}$ = (41 + 27)/(50 + 50) = .68, the test statistic value is

$z = \dfrac{(.82 - .54) - 0}{\sqrt{.68(1-.68)(1/50 + 1/50)}}$ = 3.00. The two-tailed P-value is $2P(Z \ge 3.00) = .003 < \alpha = .01$; equivalently, $|z| = |3.00| \ge z_{.005} = 2.576$. Thus, H_0 is rejected at the .01 level, and we conclude that the proportions of all students who would agree to be surveyed by Melissa and Kristine are *not* the same.

b. Not necessarily. Accent is not the only feature that makes Melissa and Kristine different (they are two different people, after all). Any other distinction between the two women serves as a competing explanation for why students were more likely to accede to Melissa than to Kristine. In an ideal study, one person would do all 100 interview attempts, randomly deciding which of two accents to present to each potential subject.

67. Using $p_1 = q_1 = p_2 = q_2 = .5$, $w = 2(1.96)\sqrt{\left(\frac{.25}{n} + \frac{.25}{n}\right)} = \frac{2.7719}{\sqrt{n}}$, so w = .1 requires n = 769.

69.

a. The "after" success probability is $p_1 + p_3$ while the "before" probability is $p_1 + p_2$, so $p_1 + p_3 > p_1 + p_2$ becomes $p_3 > p_2$; thus, we wish to test $H_0 : p_3 = p_2$ versus $H_a : p_3 > p_2$.

b. The estimator of $(p_1 + p_3) - (p_1 + p_2) = p_3 - p_2$ is $\hat{p}_3 - \hat{p}_2 = \dfrac{X_3 - X_2}{n}$.

c. When H_0 is true, $p_2 = p_3$, so $V\left(\dfrac{X_3 - X_2}{n}\right) = \dfrac{p_2 + p_3 - (p_2 - p_3)^2}{n} = \dfrac{p_2 + p_3}{n}$, which is estimated by $\dfrac{\hat{p}_2 + \hat{p}_3}{n} = \dfrac{X_2 + X_3}{n^2}$. The z statistic is then $\dfrac{\dfrac{X_3 - X_2}{n}}{\sqrt{\dfrac{X_2 + X_3}{n^2}}} = \dfrac{X_3 - X_2}{\sqrt{X_2 + X_3}}$.

d. The computed value of z is $\dfrac{200 - 150}{\sqrt{200 + 150}} = 2.68$, so P-value = 1 – Φ(2.68) = .0037. At level .01, H_0 can be rejected, but at level .001, H_0 would not be rejected.

Section 10.5

71.

a. From Table A.8, column 5, row 8, $F_{.01,5,8} = 3.69$.

b. From column 8, row 5, $F_{.01,8,5} = 4.82$.

c. $F_{.95,5,8} = \dfrac{1}{F_{.05,8,5}} = .207$.

d. $F_{.95,8,5} = \dfrac{1}{F_{.05,5,8}} = .271$

e. $F_{.01,10,12} = 4.30$

f. $F_{.99,10,12} = \dfrac{1}{F_{.01,12,10}} = \dfrac{1}{4.71} = .212$.

g. $F_{.05,6,4} = 6.16$, so $P(F \le 6.16) = .95$.

h. Since $F_{.99,10,5} = \dfrac{1}{5.64} = .177$, $P(.177 \le F \le 4.74) = P(F \le 4.74) - P(F \le .177)$
$= .95 - .01 = .94$.

73. With $\sigma_1 =$ true standard deviation for not-fused specimens and $\sigma_2 =$ true standard deviation for fused specimens, we test $H_0: \sigma_1 = \sigma_2$ v. $H_a: \sigma_1 > \sigma_2$. The calculated test statistic is $f = \dfrac{(277.3)^2}{(205.9)^2} = 1.814$. With numerator df = m – 1 = 10 – 1 = 9, and denominator df = n – 1 = 8 – 1 = 7, $f = 1.814 < 2.72 = F_{.10,9,7}$. We can say that the P-value > .10, which is obviously > .01, so we cannot reject H_0. There is not sufficient evidence that the standard deviation of the strength distribution for fused specimens is smaller than that of not-fused specimens.

75. With $\sigma_1 =$ true standard deviation for high rail breaks and $\sigma_2 =$ true standard deviation for low rail breaks, we test $H_0: \sigma_1 = \sigma_2$ v. $H_a: \sigma_1 > \sigma_2$. The calculated test statistic is $f = s_1^2 / s_2^2$ = $(145.1)^2/(69.3)^2$ = 4.38. With numerator df = m – 1 = 12 – 1 = 11 and denominator df = n – 1 = 10 – 1 = 9, we reject H_0 if $f \geq F_{.01,11,9} = 5.18$. Since 4.38 < 5.18, we fail to reject H_0 at the .01 level and cannot conclude the true sd of repair times is greater for high rail breaks. (The P-value is about .018, so H_0 *would* be rejected at the .05 significance level.)

77. From Exercise 24, $m = n = 17$, $s_1 = 4.5$ kg, and $s_2 = 3.1$ kg. With equal sample sizes, the only required critical value is $F_{.025,17-1,17-1} = 2.76$. Then a 95% CI for the ratio of population standard deviations, σ_1 / σ_2, is $\left(\dfrac{s_1}{s_2}\cdot\dfrac{1}{\sqrt{2.76}}, \dfrac{s_1}{s_2}\cdot\sqrt{2.76}\right) = \left(\dfrac{4.5}{3.1}\cdot\dfrac{1}{\sqrt{2.76}}, \dfrac{4.5}{3.1}\cdot\sqrt{2.76}\right)$ = (0.87, 2.41).

Section 10.6

79.

a. Software gives the following results: group L has a mean GPA of 3.367 with a sd of 0.514; group N has a mean GPA of 2.920 with a sd of 0.598; the estimated df is ν = 56. From these, a 95% CI for $\mu_1 - \mu_2$ is (.158, .735).

b. You can create the bootstrap distribution of differences by using the code from Chapter 8 separately on each of the two samples, then computing differences of the side-by-side pairs. Answers will vary, but the bootstrap distribution of differences looks quite normal.

c. Answers will vary; one simulation gave $s_{\text{boot}} = 0.141$. This suggests the following 95% CI for $\mu_1 - \mu_2$: (3.367 – 2.920) ± $t_{.025,56}(0.141) \approx 0.447 \pm (1.96)(0.141)$ = (.171, .723).

d. Answers will vary; choosing the 25th bootstrap value from each end of the distribution in one simulation gave a percentile interval of (.156, .740).

e. All three intervals are very close to each other, suggesting the sampling distribution of the difference of means is normal here, as noted above in (b).

f. Students on lifestyle floors appear to have a higher mean GPA, somewhere between ~.16 higher and ~.73 higher.

81.

a. The standard deviations of the two samples are 0.514 and 0.598. The relevant critical value is $F_{.025,29,29} = 2.101$. Thus, a 95% CI for $\frac{\sigma_1}{\sigma_2}$ is $\left(\frac{0.514}{0.598}\frac{1}{\sqrt{2.101}}, \frac{0.514}{0.598}\sqrt{2.101}\right) =$ (0.593,1.246). Normal probability plots of the two samples shows some noticeable departures from normality, more so that we are usually willing to accept for this F procedure.

b. The R code below assumes two vectors, L and N, contain the original data (same as Exercise 79).

```
ratio = rep(0,5000)
for (i in 1:5000){
  L.resamp = sample(L,length(L),replace=T)
  N.resamp = sample(N,length(N),replace=T)
  ratio[i] = sd(L.resamp)/sd(N.resamp)
}
```

Find the 25$^{\text{th}}$ value from either end of the ordered list of ratios. Answers will vary; one bootstrap simulation gave (0.568,1.289).

c. The intervals are fairly similar, although that won't be the case for every simulation run. Due to lack of normality, we have more faith in the bootstrap CI. Notice that both intervals contain 1, suggesting the two population standard deviations could be equal — this is consistent with the previous exercise.

83.

a. You can create the bootstrap distribution of differences by using the median code from Chapter 8 separately on each of the two samples, then computing differences of the side-by-side pairs. The bootstrap distribution of differences of medians is definitely not normal: the distribution is multimodal and positively skewed.

b. Answers will vary; one simulation gave $s_{\text{boot}} = 2.5657$. The medians of the two original samples are 13.88 and 8.47. This suggests the following 95% CI for $\tilde{\mu}_1 - \tilde{\mu}_2$: (13.88 – 8.47) ± $z_{.025}$ (2.5657) ≈ (5.41) ± (1.96)(2.5657) = (0.38, 10.44).

c. Answers will vary; choosing the 25$^{\text{th}}$ bootstrap value from each end of the distribution in one simulation gave a percentile interval of (0.4706, 10.0294).

d. The interval in (c) is slightly narrower, but neither includes zero. It is surprising that they are so close, since (b) relies on a normally distributed sampling distribution, which does not exist here.

e. The intervals from the previous exercise are considerably narrower (more "precise") than those for the difference in population medians. We can more precisely measure the difference in population means with the bootstrap in this particular case.

85.

a. For the test of H_0: $\mu_1 - \mu_2 = 0$ versus H_a: $\mu_1 - \mu_2 \neq 0$, our test statistic is $t = \frac{(10.59 - 5.71) - 0}{\sqrt{4.41^2/10 + 3.92^2/10}} = 2.62$; the estimated df is ν = 17. The 2-sided *P*-value is roughly $2P(|T| > 2.6) = 2(.009) = .018$. Hence, we reject H_0 at the α = .05 level and conclude the two population means are different. Neither of the probability plots looks very linear, but it's difficult to detect moderate deviations from normality with so few observations.

b. In the R code below, the data is read is as a data frame called df with two columns, Time and Group. The first lists the times for each rat, while the second has B and C labels.

```
N = 5000
diff = rep(0,N)
for (i in 1:N){
  resample = sample(df$Time, length(df$Time), replace=T)
  C.resamp = resample[df$Group=="C"]
  B.resamp = resample[df$Group=="B"]
  diff[i] = mean(C.resamp) - mean(B.resamp)
}
```

Run this code, then find the proportion of these differences in means that are greater than our observed difference, 10.59 – 5.71 = 4.88. Double this proportion to get the 2-sided *P*-value. Answers will vary; in one bootstrap simulation, the one-sided proportion was .0108, giving 2(.0108) = .02 as our two-sided *P*-value.

c. The answers to (a) and (b) are quite similar; in particular, both reject the null hypothesis of equal means at the $\alpha = .05$ level. This is not surprising, since the sampling distribution relevant to (a) was indeed normal (see the previous exercise).

87.

a. The standard deviations of the two samples are 3.26 and 1.54, for an *F*-ratio of $f = 4.46$. Compare this to $F_{.05,6,5} = 4.95$ and $F_{.95,6,5} = 1/F_{.05,5,6} = 1/4.39 = 0.228$: since $0.228 < 4.46 < 4.95$, we fail to reject the hypothesis that $\sigma_1 = \sigma_2$ at the $\alpha = .10$ level. The finaska barley group shows some deviation from normality, but it's difficult to detect a real departure with such a small sample.

b. In the R code below, the data is read is as a data frame called df with two columns, Barley and Gain. The first lists T's and F's, while the second has the weight gains. The entire list of weight gains is randomly permuted, then the combined sample is split according to the T and F labels. Finally, the ratio of the variances of the T and F resamples is calculated.

```
ratio = rep(0,5000)
for (i in 1:5000){
  resample = sample(df$Gain, length(df$Gain), replace=T)
  T.resamp = resample[df$Barley=="T"]
  F.resamp = resample[df$Barley=="F"]
  ratio[i] = var(T.resamp)/var(F.resamp)
}
```

The observed ratio is $3.26^2/1.54^2 = 4.48$. For one run of the above code, the proportion of ratio values that were ≥ 4.48 was .086. Double this to obtain a two-sided *P*-value 2(.086) = .172. Thus, we (again) fail to reject the null hypothesis of equal population variances (or standard deviations).

c. In either case, we have no statistically significant evidence to suggest the population standard deviations are unequal.

89.

a. Use the code provided in the solution to Exercise 85(b). The observed difference in sample means is 3.47. In one run, the proportion of differences ≥ 3.47 was .019. The resulting two-sided *P*-value is 2(.019) = .038. Thus, we reject the null hypothesis of equal population means at the $\alpha = .05$ level.

b. The result in (a) matches closely the result in Example 10.8; even the *P*-values are fairly close (.032 v .038). This comes as no surprise, since both procedures are valid: the large sample sizes permit a large-sample z-test, and the shapes of the distributions of the two samples are fairly similar (which is important for the validity of the permutation test).

91.

a. Software gives the following results: $\bar{d} = 9.126$, $s_d = 6.893$. So, a 95% confidence interval for μ_D is $9.126 \pm t_{.025,26}(6.893)/\sqrt{27}$ = (\$6.40, \$11.85). The 27 differences are grossly non-normal (heavily left-skewed); however, with a moderate sample size of n = 27, the effects of the CLT may begin to appear in the sampling distribution of $\bar{D}$.

b. Use the code provided in Chapter 8. The bootstrap distribution of $\bar{d}$ is still quite non-normal (left-skewed).

c. Answers will vary; one simulation gave $s_{\text{boot}} = 1.305$. This suggests the following 95% CI for μ_D: $9.126 \pm t_{.025,26}(1.305) = 9.126 \pm 2.056(1.305)$ = (\$6.44, \$11.81).

d. Answers will vary; choosing the 25$^{\text{th}}$ bootstrap value from each end of the distribution in one simulation gave a percentile interval of (\$6.23, \$11.51).

e. The intervals in (a) and (c) are similar; however, the interval in (d) is shifted to the left, reflecting the left-skewedness of the sampling/bootstrap distribution of $\bar{d}$. This suggests a slight problem with the symmetric intervals of (a) and (c).

f. On average, books cost between \$6.23 and \$11.51 more with Amazon than at the campus bookstore!

93. Both the bootstrap and the randomized permutation test simulate random sampling from a desired distribution in order to provide a confidence interval (bootstrap only) or to test a hypothesis (either method). The bootstrap method assumes our sample faithfully represents its population, so that sampling with replacement from the sample is equivalent to creating iid observations from the population. We then use these bootstrap samples to create a faithful representation of the sampling distribution of our relevant statistic (a sample mean or sd, a difference of two means, a median, whatever). Permutation tests are only used for comparison of two populations, and we make a different assumption: under the null hypothesis, the two populations of interest are identically distributed, and so our $m+n$ observations are really from the same distribution.

Supplementary Exercises

95. With sample sizes 56 and 59, the degrees of freedom must be at least 55 (see Exercise 97; notice we cannot estimate ν because we do not have the standard deviations). Thus, from Table A.5, $t = 6.07$ is statistically significant at any α level: .05, .01, .001. The mean number of ingredients selected by the scale-down group is indeed significantly greater than for the build-up group. This same principle might be applied to features on a new car, for example.

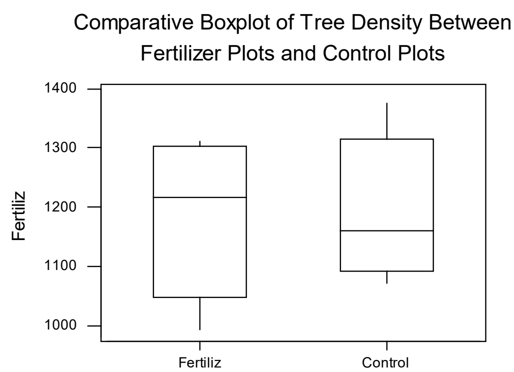
97. Since $m < n$, $\nu = \dfrac{[(se_1)^2 + (se_2)^2]^2}{(se_1)^4/(m-1) + (se_2)^4/(n-1)} > \dfrac{[(se_1)^2 + (se_2)^2]^2}{(se_1)^4/(m-1) + (se_2)^4/(m-1)} =$

$(m-1)\dfrac{[(se_1)^2 + (se_2)^2]^2}{(se_1)^4 + (se_2)^4}$; replacing n by m above increased the denominator, which decreased the overall fraction. Then, if we expand the numerator of the remaining fraction,

$\dfrac{[(se_1)^2 + (se_2)^2]^2}{(se_1)^4 + (se_2)^4} = \dfrac{(se_1)^4 + (se_2)^4 + 2(se_1)^2(se_2)^2}{(se_1)^4 + (se_2)^4} > 1$, and we conclude $\nu > (m-1)(1) = m-1$. So, a conservative estimate of the df for the 2-sample t procedures is min(m – 1, n – 1). This is easier to compute, but lowering df will result in a wider margin of error (for a CI) or less power (for a hypothesis test).

99.

a. Although the median of the fertilizer plot is higher than that of the control plots, the fertilizer plot data appears negatively skewed, while the opposite is true for the control plot data.



b. A test of $H_0: \mu_1 - \mu_2 = 0$ vs. $H_a: \mu_1 - \mu_2 \neq 0$ yields a t value of –.20 and a two-tailed P-value of .85 (df = 13). We would fail to reject H_0; the data does not indicate a significant difference in the means.

c. With 95% confidence we can say that the true average difference between the tree density of the fertilizer plots and that of the control plots is somewhere between –144 and 120. Since this interval contains 0, 0 is a plausible value for the difference, which further supports the conclusion based on the P-value.

101. The center of any confidence interval for $\mu_1 - \mu_2$ is always $\bar{x}_1 - \bar{x}_2$, so $\bar{x}_1 - \bar{x}_2 = \dfrac{-473.3 + 1691.9}{2} = 609.3$. Furthermore, half of the width of this interval is $\dfrac{1691.9 - (-473.3)}{2} = 1082.6$. Equating this value to the expression on the right of the 95% confidence interval formula, $1082.6 = (1.96)\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$, we find $\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}} = \dfrac{1082.6}{1.96} = 552.35$. For a 90% interval, the associated z value is 1.645, so the 90% confidence interval is then $609.3 \pm (1.645)(552.35) = 609.3 \pm 908.6 \;\; = (-299.3, 1517.9)$.

103. $m = n = 40$, $\bar{x} = 3975.0$, $s_1 = 245.1$, $\bar{y} = 2795.0$, $s_2 = 293.7$. The large sample 99% confidence interval for $\mu_1 - \mu_2$ is $(3975.0 - 2795.0) \pm 2.58\sqrt{\dfrac{245.1^2}{40} + \dfrac{293.7^2}{40}}$ = (1020, 1340). The value 0 is not contained in this interval so we can state that, with very high confidence, the value of $\mu_1 - \mu_2$ is not 0, which is equivalent to concluding that the population means are not equal.

105. Let μ_1 denote the true average tear length for Brand A and let μ_2 denote the true average tear length for Brand B. The relevant hypotheses are $H_0: \mu_1 - \mu_2 = 0$ vs. $H_a: \mu_1 - \mu_2 > 0$. Assuming both populations have normal distributions, the two-sample t test is appropriate. $m = 16$, $\bar{x} = 74.0$, $s_1 = 14.8$, $n = 14$, $\bar{y} = 61.0$, $s_2 = 12.5$, so the approximate df is

$$\nu = \frac{\left(\frac{14.8^2}{16} + \frac{12.5^2}{14}\right)^2}{\frac{\left(\frac{14.8^2}{16}\right)^2}{15} + \frac{\left(\frac{12.5^2}{14}\right)^2}{13}} = 27.97$$

, which we round down to 27. The test statistic is

$$t = \frac{74.0 - 61.0}{\sqrt{\frac{14.8^2}{16} + \frac{12.5^2}{14}}} \approx 2.6$$

. From Table A.7, the P-value = $P(T > 2.6)$ = .007. At a significance level of .05, H_0 is rejected, and we conclude that the average tear length for Brand A is larger than that of Brand B.

107.

a. Let μ_1 = true mean AEDI score improvement for all 2001 students. We wish to test H_0: $\mu_1 = 0$ versus H_a: $\mu_1 > 0$; the former implies no improvement, on average, while the latter implies positive average improvement. A one-sample t test is appropriate: $t = \dfrac{5.48 - 0}{13.83/\sqrt{37}} = 2.41$, and at 36 df, P-value = .011. At a 5% significance level, the data indicate statistically significant improvement in AEDI score across the semester.

b. Let μ_2 = true mean AEDI score improvement for all 2002 students. Repeating part **a**, $t = \dfrac{6.31 - 0}{13.20/\sqrt{21}}$ = 2.19 and P-value = .020 at 20 df. Again, we have evidence of a statistically significant improvement.

c. Now let's perform a *two-sample t* test. An "Enron effect" would mean that AEDI improvements were higher in 2002 than in 2001, so the hypotheses are H_0: $\mu_1 - \mu_2 = 0$ vs H_0: $\mu_1 - \mu_2 < 0$ (i.e., $\mu_2 > \mu_1$). The two-sample t statistic is $t = \dfrac{(5.48 - 6.31) - 0}{\sqrt{13.83^2/37 + 13.20^2/21}} = -$0.23. Software estimates ν = 41, and the lower-tailed P-value is .411. With such a large P-value, H_0 is definitely not rejected, and the data do *not* provide evidence of a significantly higher improvement in 2002 compared to 2001. That is, the data do not convince us of an "Enron effect."

109. Let μ_1 denote the true average ratio for young men and μ_2 denote the true average ratio for elderly men. Assuming both populations from which these samples were taken are normally distributed, the relevant hypotheses are $H_0: \mu_1 - \mu_2 = 0$ vs. $H_a: \mu_1 - \mu_2 > 0$. The value of the test statistic is $t = \dfrac{(7.47 - 6.71)}{\sqrt{\dfrac{(.22)^2}{13} + \dfrac{(.28)^2}{12}}} = 7.5$. The df = 20 and the P-value is $P(T > 7.5) \approx 0$.

Since the P-value is $< \alpha = .05$, we reject H_0. We have sufficient evidence to claim that the true average ratio for young men exceeds that for elderly men.

111. NO, since a 2-sample t test is the wrong analysis here! Instead, we should perform a paired t test. For the data provided, $\bar{d} = 0.3$, $s_D = 0.276$, and $t = 2.67$ at 5 df. This has a corresponding 2-sided P-value of 0.045, and so we reject the hypothesis of zero mean difference at the $\alpha = .05$ significance level.

113. Because of the nature of the data, we will use a paired t test. We obtain the differences by subtracting intake value from expenditure value. We are testing the hypotheses H_0: $\mu_D = 0$ vs H_a: $\mu_D \neq 0$. The test statistic $t = \dfrac{1.757}{1.197/\sqrt{7}} = 3.88$ with df $= n - 1 = 6$ leads to a P-value of $2P(T > 3.88) \approx .008$. Using either significance level .05 or .01, we would reject the null hypothesis and conclude that there is a difference between average intake and expenditure. However, at significance level .001, we would not reject.

115.

a. Let μ_1 = the true mean test validity rating under the positive feedback condition, and let μ_2 = the true mean test validity rating under the negative feedback condition. The hypotheses of interest are H_0: $\mu_1 - \mu_2 = 0$ vs H_a: $\mu_1 - \mu_2 > 0$. With such large sample sizes, H_0 will be rejected if $t > z_{.01} = 2.33$. Here, $t = \dfrac{(6.95 - 5.51) - 0}{\sqrt{1.09^2/123 + 0.79^2/123}} = 11.86$, so we clearly reject H_0. The data affirms that negative feedback is associated with a lower average validity rating than positive feedback.

b. Repeat part **a**. Now $t = \dfrac{(6.62 - 5.36) - 0}{\sqrt{1.19^2/123 + 1.00^2/123}} = 8.99$, and again we clearly reject H_0. The data verifies that students receiving positive feedback rate face-reading as more important, on average, that students receiving negative feedback.

c. Yes. Because students were *randomly assigned* to the two experimental groups, it is reasonable to conclude that the observed effects in **a** and **b** are attributable to positive vs negative feedback. All competing explanations for these significant differences should be roughly "balanced" across the two treatment groups.

117. $\Delta_0 = 0$, $\sigma_1 = \sigma_2 = 10$, $d = 1$, $\sigma = \sqrt{\dfrac{200}{n}} = \dfrac{14.142}{\sqrt{n}}$, so $\beta = \Phi\left(1.645 - \dfrac{\sqrt{n}}{14.142}\right)$, giving β = .9015, .8264, .0294, and .0000 for n = 25, 100, 2500, and 10,000 respectively. If the μ_i's referred to true average IQs resulting from two different conditions, $\mu_1 - \mu_2 = 1$ would have little practical significance, yet very large sample sizes would yield statistical significance in this situation.

119. $H_0: p_1 = p_2$ will be rejected at level α in favor of $H_a: p_1 > p_2$ if $z \geq z_\alpha$. With $\hat{p}_1 = \frac{250}{2500} = .10$ and $\hat{p}_2 = \frac{167}{2500} = .0668$, $\hat{p} = .0834$ and $z = \dfrac{.0332}{.0079} = 4.2$, so H_0 is rejected at any reasonable α level. It appears that a response is more likely for a white name than for a black name.

121. First, $V\left(\overline{X}-\overline{Y}\right)=\frac{\mu_1}{m}+\frac{\mu_2}{n}=\mu\left(\frac{1}{m}+\frac{1}{n}\right)$ when H_0 is true, where μ can be estimated for the variance by the pooled estimate $\hat{\mu}_{pooled}=\frac{m\overline{X}+n\overline{Y}}{m+n}$. With the obvious point estimates $\hat{\mu}_1=\overline{X}$ and $\hat{\mu}_2=\overline{Y}$, we have a large-sample test statistic of $Z=\frac{(\overline{X}-\overline{Y})-0}{\sqrt{\hat{\mu}_{pooled}\left(\frac{1}{m}+\frac{1}{n}\right)}}=\frac{\overline{X}-\overline{Y}}{\sqrt{\frac{\overline{X}}{n}+\frac{\overline{Y}}{m}}}$.

With $\bar{x}=1.616$ and $\bar{y}=2.557$, $z=-5.3$ and P-value $=P(|Z|\geq|-5.3|)=2\Phi(-5.3)\approx 0$, so we would certainly reject H_0: $\mu_1=\mu_2$ in favor of H_a: $\mu_1\neq\mu_2$.

123. Define standard normal and chi-squared rvs as follows: $Z=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\sigma^2/m+\sigma^2/n}}$, $W=(m+n-2)S_p^2/\sigma^2$ (df $=m+n-2$). Then, by definition, the rv $\frac{Z+\delta}{\sqrt{W/\text{df}}}$ has, by definition, a noncentral t distribution. Substitute Z and W above along with δ specified in the exercise; we hope to show the result is T_p. Along the way, note that Δ' is just shorthand notation for $(\mu_1-\mu_2)$.

$$\frac{Z+\delta}{\sqrt{W/\text{df}}}=\frac{\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\sigma^2/m+\sigma^2/n}}+\frac{\Delta'-\Delta_0}{\sigma\sqrt{1/m+1/n}}}{\sqrt{(m+n-2)S_p^2/\sigma^2/(m+n-2)}}=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)+\Delta'-\Delta_0}{\sigma\sqrt{1/m+1/n}}\cdot\frac{1}{\sqrt{S_p^2/\sigma^2}}$$

$$=\frac{(\overline{X}-\overline{Y})-\Delta_0}{S_p\sqrt{1/m+1/n}}=T_p$$