CHAPTER 9

Section 9.1

1.

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

b. $V(\overline{X}-\overline{Y}) = V(\overline{X}) + V(\overline{Y}) = \dfrac{s_1^{\ 2}}{m} + \dfrac{s_2^{\ 2}}{n} = \dfrac{(1.8)^2}{100} + \dfrac{(2.0)^2}{100} = .0724$, and the s.d. of $\overline{X}-\overline{Y} = \sqrt{.0724} = .2691$.

c. A normal curve with mean and s.d. 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.

2. The test statistic value is $z = \dfrac{\bar{x}-\bar{y}}{\sqrt{\dfrac{s_1^2}{m}+\dfrac{s_2^2}{n}}}$, and H_o will be rejected if either $z \geq 1.96$ or $z \leq -1.96$. We compute $z = \dfrac{42{,}500 - 40{,}400}{\sqrt{\dfrac{2200^2}{45}+\dfrac{1900^2}{45}}} = \dfrac{2100}{433.33} = 4.85$. Since 4.85 > 1.96, reject H_o and conclude that the two brands differ with respect to true average tread lives.

3. The test statistic value is $z = \dfrac{(\bar{x}-\bar{y}) - 5000}{\sqrt{\dfrac{s_1^2}{m}+\dfrac{s_2^2}{n}}}$, and H_o will be rejected at level .01 if $z \geq 2.33$. We compute $z = \dfrac{(43{,}500 - 36{,}800) - 5000}{\sqrt{\dfrac{2200^2}{45}+\dfrac{1500^2}{45}}} = \dfrac{700}{396.93} = 1.76$, which is not > 2.33, so we don't reject H_o and conclude that the true average life for radials does not exceed that for economy brand by more than 500.

4.

a. From Exercise 2, the C.I. is

$$(\bar{x}-\bar{y})\pm(1.96)\sqrt{\frac{s_1^2}{m}+\frac{s_2^2}{n}}=2100\pm 1.96(433.33)=2100\pm 849.33$$

$=(1250.67, 2949.33)$. In the context of this problem situation, the interval is moderately wide (a consequence of the standard deviations being large), so the information about m_1 and m_2 is not as precise as might be desirable.

b. From Exercise 3, the upper bound is $5700+1.645(396.93)=5700+652.95=6352.95$.

5.

a. H_a says that the average calorie output for sufferers is more than 1 cal/cm^2/min below that for nonsufferers. $\sqrt{\frac{s_1^2}{m}+\frac{s_2^2}{n}}=\sqrt{\frac{(.04)^2}{10}+\frac{(.16)^2}{10}}=.1414$, so $z=\frac{(.64-2.05)-(-1)}{.1414}=-2.90$. At level .01, H_o is rejected if $z\le -2.33$; since –2.90 < -2.33, reject H_o.

b. $P=\Phi(-2.90)=.0019$

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

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

6.

a. H_o should be rejected if $z \geq 2.33$. Since $z = \dfrac{(18.12 - 16.87)}{\sqrt{\dfrac{2.56}{40} + \dfrac{1.96}{32}}} = 3.53 \geq 2.33$, H_o should be rejected at level .01.

b. $b(1) = \Phi\left(2.33 - \dfrac{1-0}{.3539}\right) = \Phi(-.50) = .3085$

c. $\dfrac{2.56}{40} + \dfrac{1.96}{n} = \dfrac{1}{(1.645 + 1.28)^2} = .1169 \Rightarrow \dfrac{1.96}{n} = .0529 \Rightarrow n = 37.06$, so use n = 38.

d. Since n = 32 is not a large sample, it would no longer be appropriate to use the large sample test. A small sample t procedure should be used (section 9.2), and the appropriate conclusion would follow.

7.

1 Parameter of interest: $m_1 - m_2 =$ the true difference of means for males and females on the Boredom Proneness Rating. Let $m_1 =$ men's average and $m_2 =$ women's average.

2 H_o: $m_1 - m_2 = 0$

3 H_a: $m_1 - m_2 > 0$

4 $z = \dfrac{(\bar{x} - \bar{y}) - \Delta_o}{\sqrt{\dfrac{s_1^2}{m} + \dfrac{s_2^2}{n}}} = \dfrac{(\bar{x} - \bar{y}) - 0}{\sqrt{\dfrac{s_1^2}{m} + \dfrac{s_2^2}{n}}}$

5 RR: $z \geq 1.645$

6 $z = \dfrac{(10.40 - 9.26) - \Delta_o}{\sqrt{\dfrac{4.83^2}{97} + \dfrac{4.68^2}{148}}} = 1.83$

7 Reject H_o. The data indicates the Boredom Proneness Rating is higher for males than for females.

8.

a.

1 Parameter of interest: $m_1 - m_2 =$ the true difference of mean tensile strength of the 1064 grade and the 1078 grade wire rod. Let $m_1 =$ 1064 grade average and $m_2 =$ 1078 grade average.

2 H$_o$: $m_1 - m_2 = -10$

3 H$_a$: $m_1 - m_2 < -10$

4 $$z = \frac{(\bar{x} - \bar{y}) - \Delta_o}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}} = \frac{(\bar{x} - \bar{y}) - (-10)}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}}$$

5 RR: $p - value < a$

6 $$z = \frac{(107.6 - 123.6) - (-10)}{\sqrt{\frac{1.3^2}{129} + \frac{2.0^2}{129}}} = \frac{-6}{.210} = -28.57$$

7 For a lower-tailed test, the p-value = $\Phi(-28.57) \approx 0$, which is less than any a , so reject H$_o$. There is very compelling evidence that the mean tensile strength of the 1078 grade exceeds that of the 1064 grade by more than 10.

b. The requested information can be provided by a 95% confidence interval for $m_1 - m_2$:

$$(\bar{x} - \bar{y}) \pm 1.96\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}} = (-6) \pm 1.96(.210) = (-6.412, -5.588) .$$

9.

a. point estimate $\bar{x} - \bar{y} = 19.9 - 13.7 = 6.2$. It appears that there could be a difference.

b.

H$_o$: $m_1 - m_2 = 0$,H$_a$: $m_1 - m_2 \neq 0$, $z = \dfrac{(19.9 - 13.7)}{\sqrt{\frac{39.1^2}{60} + \frac{15.8^2}{60}}} = \dfrac{6.2}{5.44} = 1.14$, and the p-value = 2[P(z > 1.14)] = 2(.1271) = .2542. The p value is larger than any reasonable α, so we do not reject H$_0$. There is no significant difference.

c. No. With a normal distribution, we would expect most of the data to be within 2 standard deviations of the mean, and the distribution should be symmetric. 2 sd's above the mean is 98.1, but the distribution stops at zero on the left. The distribution is positively skewed.

d. We will calculate a 95% confidence interval for μ, the true average length of stays for patients given the treatment. $19.9 \pm 1.96\dfrac{39.1}{\sqrt{60}} = 19.9 \pm 9.9 = (10.0, 21.8)$

10.

a. The hypotheses are H$_o$: $\mu_1 - \mu_2 = 5$ and H$_a$: $\mu_1 - \mu_2 > 5$. At level .001, H$_o$ should be rejected if $z \geq 3.08$. Since $z = \dfrac{(65.6 - 59.8) - 5}{.2272} = 2.89 < 3.08$, H$_o$ cannot be rejected in favor of H$_a$ at this level, so the use of the high purity steel cannot be justified.

b. $\mu_1 - \mu_2 - \Delta_o = 1$, so $\beta = \Phi\left(3.08 - \dfrac{1}{.2272}\right) = \Phi(-.53) = .2891$

11. $(\overline{X} - \overline{Y}) \pm z_{\alpha/2}\sqrt{\dfrac{\sigma_1^2}{m} + \dfrac{\sigma_2^2}{n}}$. Standard error $= \dfrac{\sigma}{\sqrt{n}}$. Substitution yields $(\overline{x} - \overline{y}) \pm z_{\alpha/2}\sqrt{(SE_1)^2 + (SE_2)^2}$. Using $\alpha = .05$, $z_{\alpha/2} = 1.96$, so $(5.5 - 3.8) \pm 1.96\sqrt{(0.3)^2 + (0.2)^2} = (0.99, 2.41)$. Because we selected $\alpha = .05$, we can state that when using this method with repeated sampling, the interval calculated will bracket the true difference 95% of the time. The interval is fairly narrow, indicating precision of the estimate.

12. The C.I. is $(\overline{x} - \overline{y}) \pm 2.58\sqrt{\dfrac{s_1^2}{m} + \dfrac{s_2^2}{n}} = (-8.77) \pm 2.58\sqrt{.9104} = -8.77 \pm 2.46$ $= (-11.23, -6.31)$. With 99% confidence we may say that the true difference between the average 7-day and 28-day strengths is between -11.23 and -6.31 N/mm^2.

13. $\sigma_1 = \sigma_2 = .05$, d = .04, $\alpha = .01$, $\beta = .05$, and the test is one-tailed, so

$$n = \frac{(.0025 + .0025)(2.33 + 1.645)^2}{.0016} = 49.38$$

, so use n = 50.

14. The appropriate hypotheses are H$_o$: $\theta = 0$ vs. H$_a$: $\theta < 0$, where $\theta = 2\mu_1 - \mu_2$. ($\theta < 0$ is equivalent to $2\mu_1 < \mu_2$, so normal is more than twice schizo) The estimator of θ is $\hat{\theta} = 2\overline{X} - \overline{Y}$, with $Var(\hat{\theta}) = 4Var(\overline{X}) + Var(\overline{Y}) = \dfrac{4\sigma_1^2}{m} + \dfrac{\sigma_2^2}{n}$, σ_θ is the square root of $Var(\hat{\theta})$, and $\hat{s}_\theta$ is obtained by replacing each σ_i^2 with S_i^2. The test statistic is then $\dfrac{\hat{\theta}}{s_{\hat{\theta}}}$ (since $\theta_o = 0$), and H$_o$ is rejected if $z \leq -2.33$. With $\hat{\theta} = 2(2.69) - 6.35 = -.97$ and $\hat{s}_\theta = \sqrt{\dfrac{4(2.3)^2}{43} + \dfrac{(4.03)^2}{45}} = .9236$, $z = \dfrac{-.97}{.9236} = -1.05$; Because $-1.05 > -2.33$, H$_o$ is not rejected.

15.

a. As either m or n increases, s decreases, so $\frac{m_1 - m_2 - \Delta_o}{s}$ increases (the numerator is positive), so $\left(z_a - \frac{m_1 - m_2 - \Delta_o}{s} \right)$ decreases, so $b = \Phi\left(z_a - \frac{m_1 - m_2 - \Delta_o}{s} \right)$ decreases.

b. As b decreases, z_b increases, and since z_b is the numerator of n , n increases also.

16. $z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{n}}} = \frac{.2}{\sqrt{\frac{2}{n}}}$. For n = 100, z = 1.41 and p-value = $2[1 - \Phi(1.41)] = .1586$.

For n = 400, z = 2.83 and p-value = .0046. From a practical point of view, the closeness of $\bar{x}$ and $\bar{y}$ suggests that there is essentially no difference between true average fracture toughness for type I and type I steels. The very small difference in sample averages has been magnified by the large sample sizes – statistical rather than practical significance. The p-value by itself would not have conveyed this message.

Section 9.2

17.

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

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

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

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

18. With H_o: $m_1 - m_2 = 0$ vs. H_a: $m_1 - m_2 \neq 0$, we will reject H_o if $p - value < a$.

$$n = \frac{\left(\frac{.164^2}{6} + \frac{.240^2}{5}\right)^2}{\frac{\left(\frac{.164^2}{6}\right)^2}{5} + \frac{\left(\frac{.240^2}{5}\right)^2}{4}} = 6.8 \approx 6$$

, and the test statistic

$$t = \frac{22.73 - 21.95}{\sqrt{\frac{.164^2}{6} + \frac{.240^2}{5}}} = \frac{.78}{.1265} = 6.17$$

leads to a p-value of 2[P(t > 6.17)] < 2(.0005) =.001, which is less than most reasonable $a's$, so we reject H_o and conclude that there is a difference in the densities of the two brick types.

19. For the given hypotheses, the test statistic $t = \frac{115.7 - 129.3 + 10}{\sqrt{\frac{5.03^2}{6} + \frac{5.38^2}{6}}} = \frac{-3.6}{3.007} = -1.20$, and the d.f. is

$$n = \frac{(4.2168 + 4.8241)^2}{\frac{(4.2168)^2}{5} + \frac{(4.8241)^2}{5}} = 9.96$$

, so use d.f. = 9. We will reject H_o if $t \leq -t_{.01,9} = -2.764$; since –1.20 > -2.764, we don't reject H_o.

20. We want a 95% confidence interval for $m_1 - m_2$. $t_{.025,9} = 2.262$, so the interval is $-3.6 \pm 2.262(3.007) = (-10.40, 3.20)$. Because the interval is so wide, it does not appear that precise information is available.

21. Let $m_1 =$ the true average gap detection threshold for normal subjects, and $m_2 =$ the corresponding value for CTS subjects. The relevant hypotheses are H_o: $m_1 - m_2 = 0$ vs. H_a: $m_1 - m_2 < 0$, and the test statistic $t = \frac{1.71 - 2.53}{\sqrt{.0351125 + .07569}} = \frac{-.82}{.3329} = -2.46$.

Using d.f.

$$n = \frac{(.0351125 + .07569)^2}{\frac{(.0351125)^2}{7} + \frac{(.07569)^2}{9}} = 15.1$$

, or 15, the rejection region is $t \leq -t_{.01,15} = -2.602$. Since –2.46 is not ≤ -2.602, we fail to reject H_o. We have insufficient evidence to claim that the true average gap detection threshold for CTS subjects exceeds that for normal subjects.

22. Let $m_1 =$ the true average strength for wire-brushing preparation and let $m_2 =$ the average strength for hand-chisel preparation. Since we are concerned about any possible difference between the two means, a two-sided test is appropriate. We test $H_0 : m_1 - m_2 = 0$ vs. $H_a : m_1 - m_2 \neq 0$. We need the degrees of freedom to find the rejection region:

$$n = \frac{\left(\frac{1.58^2}{12} + \frac{4.01^2}{12}\right)^2}{\frac{\left(\frac{1.58^2}{12}\right)^2}{11} + \frac{\left(\frac{4.01^2}{5}\right)^2}{11}} = \frac{2.3964}{.0039 + .1632} = 14.33$$

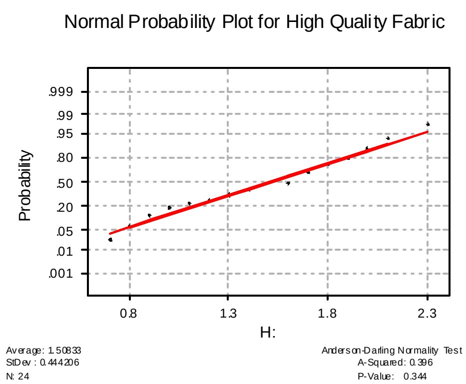
, which we round down to 14, so we reject H_o if $|t| \geq t_{.025,14} = 2.145$. The test statistic is

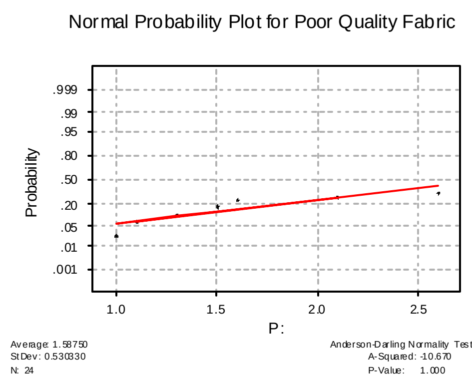
$$t = \frac{19.20 - 23.13}{\sqrt{\left(\frac{1.58^2}{12} + \frac{4.01^2}{12}\right)}} = \frac{-3.93}{1.2442} = -3.159$$

, which is ≤ -2.145, so we reject H_o and conclude that there does appear to be a difference between the two population average strengths.

23.

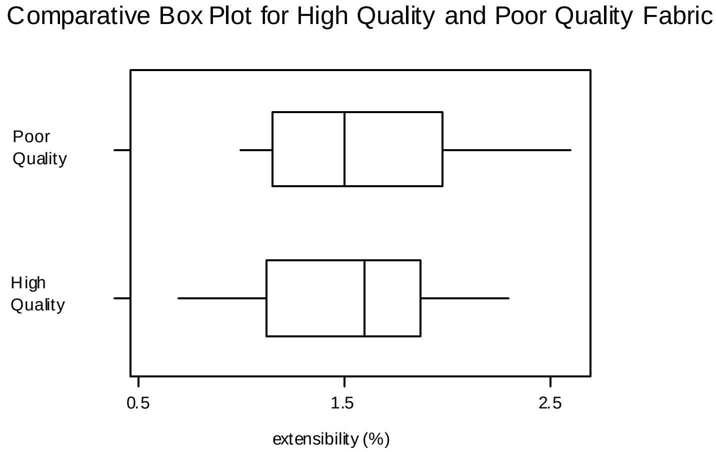
a. Normal plots





Using Minitab to generate normal probability plots, 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.

c. We test $H_0 : m_1 - m_2 = 0$ vs. $H_a : m_1 - m_2 \neq 0$. With degrees of freedom $n = \dfrac{(.0433265)^2}{.00017906} = 10.5$, which we round down to 10, and using significance level .05 (not specified in the problem), we reject H_o 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_o. There is insufficient evidence to claim that the true average extensibility differs for the two types of fabrics.

24. A 95% confidence interval for the difference between the true firmness of zero-day apples and the true firmness of 20-day apples is $(8.74 - 4.96) \pm t_{.025,n}\sqrt{\dfrac{.66^2}{20} + \dfrac{.39^2}{20}}$. We calculate the degrees of freedom $n = \dfrac{\left(\dfrac{.66^2}{20} + \dfrac{.39^2}{20}\right)^2}{\dfrac{\left(\frac{.66^2}{20}\right)^2}{19} + \dfrac{\left(\frac{.39^2}{20}\right)^2}{19}} = 30.83$, so we use 30 df, and $t_{.025,30} = 2.042$, so the interval is $3.78 \pm 2.042(.17142) = (3.43, 4.13)$. Thus, with 95% confidence, we can say that the true average firmness for zero-day apples exceeds that of 20-day apples by between 3.43 and 4.13 N.

25. We calculate the degrees of freedom $n = \dfrac{\left(\frac{5.5^2}{28}+\frac{7.8^2}{31}\right)^2}{\frac{\left(\frac{5.5^2}{28}\right)^2}{27}+\frac{\left(\frac{7.8^2}{31}\right)^2}{30}} = 53.95$, or about 54 (normally we would round down to 53, but this number is very close to 54 – of course for this large number of df, using either 53 or 54 won't make much difference in the critical t value) so the desired confidence interval is $(91.5-88.3)\pm 1.68\sqrt{\frac{5.5^2}{28}+\frac{7.8^2}{31}}$ $= 3.2 \pm 2.931 = (.269, 6.131)$. Because 0 does not lie inside this interval, we can be reasonably certain that the true difference $m_1 - m_2$ is not 0 and, therefore, that the two population means are not equal. For a 95% interval, the t value increases to about 2.01 or so, which results in the interval 3.2 ± 3.506. Since this interval does contain 0, we can no longer conclude that the means are different if we use a 95% confidence interval.

26. Let $m_1 =$ the true average potential drop for alloy connections and let $m_2 =$ the true average potential drop for EC connections. Since we are interested in whether the potential drop is higher for alloy connections, an upper tailed test is appropriate. We test $H_0 : m_1 - m_2 = 0$ vs. $H_a : m_1 - m_2 > 0$. Using the SAS output provided, the test statistic, when assuming unequal variances, is t = 3.6362, the corresponding df is 37.5, and the p-value for our upper tailed test would be ½ (two-tailed p-value) = $\frac{1}{2}(.0008) = .0004$. Our p-value of .0004 is less than the significance level of .01, so we reject H_o. We have sufficient evidence to claim that the true average potential drop for alloy connections is higher than that for EC connections.

27. The approximate degrees of freedom for this estimate are

$$n = \frac{\left(\frac{11.3^2}{6}+\frac{8.3^2}{8}\right)^2}{\frac{\left(\frac{11.3^2}{6}\right)^2}{5}+\frac{\left(\frac{8.3^2}{8}\right)^2}{7}} = \frac{893.59}{101.175} = 8.83$$, which we round down to 8, so $t_{.025,8} = 2.306$ and the desired interval is $(40.3-21.4)\pm 2.306\sqrt{\frac{11.3^2}{6}+\frac{8.3^2}{8}} = 18.9 \pm 2.306(5.4674)$ $= 18.9 \pm 12.607 = (6.3, 31.5)$. Because 0 is not contained in this interval, there is strong evidence that $m_1 - m_2$ is not 0; i.e., we can conclude that the population means are not equal. Calculating a confidence interval for $m_2 - m_1$ would change only the order of subtraction of the sample means, but the standard error calculation would give the same result as before. Therefore, the 95% interval estimate of $m_2 - m_1$ would be (-31.5, -6.3), just the negatives of the endpoints of the original interval. Since 0 is not in this interval, we reach exactly the same conclusion as before; the population means are not equal.

28. We will test the hypotheses: $H_0 : \mu_1 - \mu_2 = 10$ vs. $H_a : \mu_1 - \mu_2 > 10$. The test statistic is $t = \dfrac{(\bar{x} - \bar{y}) - 10}{\sqrt{\left(\frac{2.75^2}{10} + \frac{4.44^2}{5}\right)}} = \dfrac{4.5}{2.17} = 2.08$ The degrees of freedom

$$\nu = \frac{\left(\frac{2.75^2}{10} + \frac{4.44^2}{5}\right)^2}{\frac{\left(\frac{2.75^2}{10}\right)^2}{9} + \frac{\left(\frac{4.44^2}{5}\right)^2}{4}} = \frac{22.08}{3.95} = 5.59 \approx 6$$ and the p-value from table A.8 is approx .04, 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.

29. Let $\mu_1 =$ the true average compression strength for strawberry drink and let $\mu_2 =$ the true average compression strength for cola. A lower tailed test is appropriate. We test $H_0 : \mu_1 - \mu_2 = 0$ vs. $H_a : \mu_1 - \mu_2 < 0$. The test statistic is

$$t = \frac{-14}{\sqrt{29.4 + 15}} = -2.10 \,.\; \nu = \frac{(44.4)^2}{\frac{(29.4)^2}{14} + \frac{(15)^2}{14}} = \frac{1971.36}{77.8114} = 25.3$$, so use df=25.

The p-value $\approx P(t < -2.10) = .023$. This p-value indicates strong support for the alternative hypothesis. The data does suggest that the extra carbonation of cola results in a higher average compression strength.

30.

a. We desire a 99% confidence interval. First we calculate the degrees of freedom:

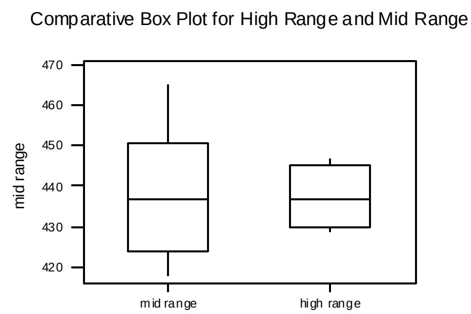
$$\nu = \frac{\left(\frac{2.2^2}{26} + \frac{4.3^2}{26}\right)^2}{\frac{\left(\frac{2.2^2}{26}\right)^2}{26} + \frac{\left(\frac{4.3^2}{26}\right)^2}{26}} = 37.24$$, which we would round down to 37, except that there is no df = 37 row in Table A.5. Using 36 degrees of freedom (a more conservative choice), $t_{.005,36} = 2.719$, and the 99% C.I. is

$$(33.4 - 42.8) \pm 2.719\sqrt{\frac{2.2^2}{26} + \frac{4.3^2}{26}} = -9.4 \pm 2.576 = (-11.98, -6.83)$$. We are very confident that the true average load for carbon beams exceeds that for fiberglass beams by between 6.83 and 11.98 kN.

b. The upper limit of the interval in part **a** does not give a 99% upper confidence bound. The 99% upper bound would be $-9.4 + 2.434(.9473) = -7.09$, meaning that the true average load for carbon beams exceeds that for fiberglass beams by at least 7.09 kN.

31.

a.



The most notable feature of these boxplots is the larger amount of variation present in the mid-range data compared to the high-range data. Otherwise, both look reasonably symmetric with no outliers present.

b. Using df = 23, a 95% confidence interval for $m_{mid-range} - m_{high-range}$ is $(438.3 - 437.45) \pm 2.069\sqrt{\frac{15.1^2}{17} + \frac{6.83^2}{11}} = .85 \pm 8.69 = (-7.84, 9.54)$. Since plausible values for $m_{mid-range} - m_{high-range}$ are both positive and negative (i.e., the interval spans zero) we would conclude that there is not sufficient evidence to suggest that the average value for mid-range and the average value for high-range differ.

32. Let $m_1 =$ the true average proportional stress limit for red oak and let $m_2 =$ the true average proportional stress limit for Douglas fir. We test $H_0 : m_1 - m_2 = 1$ vs. $H_a : m_1 - m_2 > 1$. The test statistic is $t = \frac{(8.48 - 6.65) - 1}{\sqrt{\frac{.79^2}{14} + \frac{1.28^2}{10}}} = \frac{1.83}{\sqrt{.2084}} 1.818$. With degrees of freedom

$$n = \frac{(.2084)^2}{\frac{\left(\frac{.79^2}{14}\right)^2}{13} + \frac{\left(\frac{1.28^2}{10}\right)^2}{9}} = 13.85 \approx 14$$

, the p-value $\approx P(t > 1.8) = .046$. This p-value indicates strong support for the alternative hypothesis since we would reject H_o at significance levels greater than .046. 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.

33. Let $m_1 =$ the true average weight gain for steroid treatment and let $m_2 =$ the true average weight gain for the population not treated with steroids. The exercise asks if we can conclude that m_2 exceeds m_1 by more than 5 g., which we can restate in the equivalent form: $m_1 - m_2 < -5$. Therefore, we conduct a lower-tailed test of $H_0 : m_1 - m_2 = -5$ vs. $H_a : m_1 - m_2 < -5$. The test statistic is $t = \frac{(\bar{x} - \bar{y}) - (\Delta)}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}} = \frac{32.8 - 40.5 - (-5)}{\sqrt{\frac{2.6^2}{8} + \frac{2.5^2}{10}}} = \frac{-2.7}{1.2124} = -2.23 \approx 2.2$. The approximate d.f. is $n = \frac{\left(\frac{2.6^2}{8} + \frac{2.5^2}{10}\right)^2}{\frac{\left(\frac{2.6^2}{8}\right)^2}{7} + \frac{\left(\frac{2.5^2}{10}\right)^2}{9}} = \frac{2.1609}{.1454} = 14.876$, which we round down to 14. The p-value for a lower tailed test is P(t < -2.2) = P(t > 2.2) = .022. Since this p-value is larger than the specified significance level .01, we cannot reject H_o. Therefore, this data does not support the belief that average weight gain for the control group exceeds that of the steroid group by more than 5 g.

34.

a. Following the usual format for most confidence intervals: *statistic ± (critical value)(standard error),* a pooled variance confidence interval for the difference between two means is $(\bar{x} - \bar{y}) \pm t_{a/2, m+n-2} \cdot s_p \sqrt{\frac{1}{m} + \frac{1}{n}}$.

b. The sample means and standard deviations of the two samples are $\bar{x} = 13.90$, $s_1 = 1.225$, $\bar{y} = 12.20$, $s_2 = 1.010$. The pooled variance estimate is $s_p^2 = \left(\frac{m-1}{m+n-2}\right)s_1^2 + \left(\frac{n-1}{m+n-2}\right)s_2^2 = \left(\frac{4-1}{4+4-2}\right)(1.225)^2 + \left(\frac{4-1}{4+4-2}\right)(1.010)^2 = 1.260$, so $s_p = 1.1227$. With df = m+n-1 = 6 for this interval, $t_{.025,6} = 2.447$ and the desired interval is $(13.90 - 12.20) \pm (2.447)(1.1227)\sqrt{\frac{1}{4} + \frac{1}{4}} = 1.7 \pm 1.943 = (-.24, 3.64)$. This interval contains 0, so it does not support the conclusion that the two population means are different.

c. Using the two-sample t interval discussed earlier, we use the CI as follows: First, we need to calculate the degrees of freedom. $n = \frac{\left(\frac{1.225^2}{4} + \frac{1.01^2}{4}\right)^2}{\frac{\left(\frac{1.225^2}{4}\right)^2}{3} + \frac{\left(\frac{1.01^2}{4}\right)^2}{3}} = \frac{.6302}{.0686} = 9.19 \approx 9$ so $t_{.025,9} = 2.262$. Then the interval is $(13.9 - 12.2) \pm 2.262\sqrt{\frac{1.225^2}{4} + \frac{1.01^2}{4}} = 1.70 \pm 2.262(.7938) = (-.10, 3.50)$. This interval is slightly smaller, but it still supports the same conclusion.

35. There are two changes that must be made to the procedure we currently use. First, the equation used to compute the value of the t test statistic is: $t = \dfrac{(\bar{x}-\bar{y})-(\Delta)}{s_p\sqrt{\dfrac{1}{m}+\dfrac{1}{n}}}$ where s_p is defined as in Exercise 34 above. Second, the degrees of freedom = m + n – 2. Assuming equal variances in the situation from Exercise 33, we calculate s_p as follows:

$s_p = \sqrt{\left(\dfrac{7}{16}\right)(2.6)^2 + \left(\dfrac{9}{16}\right)(2.5)^2} = 2.544$. The value of the test statistic is, then,

$t = \dfrac{(32.8-40.5)-(-5)}{2.544\sqrt{\dfrac{1}{8}+\dfrac{1}{10}}} = -2.24 \approx -2.2$. The degrees of freedom = 16, and the p-value is P (t < -2.2) = .021. Since .021 > .01, we fail to reject H_o. This is the same conclusion reached in Exercise 33.

Section 9.3

36. $\bar{d} = 7.25$, $s_D = 11.8628$

1 Parameter of Interest: $m_D =$ true average difference of breaking load for fabric in unabraded or abraded condition.

2 $H_0 : m_D = 0$

3 $H_a : m_D > 0$

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

5 RR: $t \ge t_{.01,7} = 2.998$

6 $t = \dfrac{7.25-0}{11.8628/\sqrt{8}} = 1.73$

7 Fail to reject H_o. The data does not indicate a difference in breaking load for the two fabric load conditions.

37.

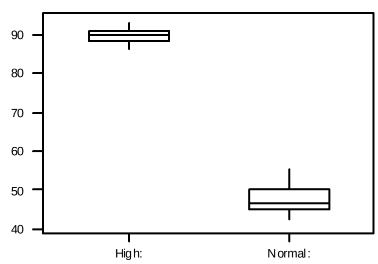
a. This exercise calls for paired analysis. First, compute the difference between indoor and outdoor concentrations of hexavalent chromium for each of the 33 houses. These 33 differences are summarized as follows: n = 33, $\bar{d} = -.4239$, $s_d = .3868$, where d = (indoor value – outdoor value). Then $t_{.025,32} = 2.037$, and a 95% confidence interval for the population mean difference between indoor and outdoor concentration is $-.4239 \pm (2.037)\left(\frac{.3868}{\sqrt{33}}\right) = -.4239 \pm .13715 = (-.5611, -.2868)$. We can be highly confident, at the 95% confidence level, that the true average concentration of hexavalent chromium outdoors exceeds the true average concentration indoors by between .2868 and .5611 nanograms/m^3.

b. A 95% prediction interval for the difference in concentration for the 34th house is $\bar{d} \pm t_{.025,32}\left(s_d\sqrt{1+\frac{1}{n}}\right) = -.4239 \pm (2.037)\left(.3868\sqrt{1+\frac{1}{33}}\right) = (-1.224, .3758)$. This prediction interval means that the indoor concentration may exceed the outdoor concentration by as much as .3758 nanograms/m^3 and that the outdoor concentration may exceed the indoor concentration by a much as 1.224 nanograms/m^3, for the 34th house. Clearly, this is a wide prediction interval, largely because of the amount of variation in the differences.

38.

a. The median of the "Normal" data is 46.80 and the upper and lower quartiles are 45.55 and 49.55, which yields an IQR of 49.55 – 45.55 = 4.00. The median of the "High" data is 90.1 and the upper and lower quartiles are 88.55 and 90.95, which yields an IQR of 90.95 – 88.55 = 2.40. The most significant feature of these boxplots is the fact that their locations (medians) are far apart.

Comparative Boxplots
for Normal and High Strength Concrete Mix



b. This data is paired because the two measurements are taken for each of 15 test conditions. Therefore, we have to work with the differences of the two samples. A quantile of the 15 differences shows that the data follows (approximately) a straight line, indicating that it is reasonable to assume that the differences follow a normal distribution. Taking differences in the order "Normal" – "High" , we find $\bar{d} = -42.23$, and $s_d = 4.34$.

With $t_{.025,14} = 2.145$, a 95% confidence interval for the difference between the population means is

$-42.23 \pm (2.145)\left(\frac{4.34}{\sqrt{15}}\right) = -42.23 \pm 2.404 = (-44.63, -39.83)$. Because 0 is not contained in this interval, we can conclude that the difference between the population means is not 0; i.e., we conclude that the two population means are not equal.

39.

a. A normal probability plot shows that the data could easily follow a normal distribution.

b. We test $H_0 : m_d = 0$ vs. $H_a : m_d \neq 0$, with test statistic

$t = \frac{\bar{d} - 0}{s_D / \sqrt{n}} = \frac{167.2 - 0}{228 / \sqrt{14}} = 2.74 \approx 2.7$. The two-tailed p-value is 2[P(t > 2.7)] = 2[.009] = .018. Since .018 < .05, we can reject H_o. There is strong evidence to support the claim that the true average difference between intake values measured by the two methods is not 0. There is a difference between them.

40.

a. H_o will be rejected in favor of H_a if either $t \geq t_{.005,15} = 2.947$ or $t \leq -2.947$. The summary quantities are $\bar{d} = -.544$, and $s_d = .714$, so $t = \frac{-.544}{.1785} = -3.05$.

Because $-3.05 \leq -2.947$, H_o is rejected in favor of H_a.

b. $s_p^2 = 7.31$, $s_p = 2.70$, and $t = \frac{-.544}{.96} = -.57$, which is clearly insignificant; the incorrect analysis yields an inappropriate conclusion.

41. We test $H_0 : m_d = 0$ vs. $H_a : m_d > 0$. With $\bar{d} = 7.600$, and $s_d = 4.178$,

$t = \frac{7.600 - 5}{4.178 / \sqrt{9}} = \frac{2.6}{1.39} = 1.87 \approx 1.9$. With degrees of freedom n – 1 = 8, the corresponding p-value is P(t > 1.9) = .047. We would reject H_o at any alpha level greater than .047. So, at the typical significance level of .05, we would (barely) reject H_o, and conclude that the data indicates that the higher level of illumination yields a decrease of more than 5 seconds in true average task completion time.

42.

1 Parameter of interest: m_d denotes the true average difference of spatial ability in brothers exposed to DES and brothers not exposed to DES. Let $m_d = m_{\exp osed} - m_{un\exp osed}$.

2 $H_0 : m_D = 0$

3 $H_a : m_D < 0$

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

5 RR: P-value < .05, df = 8

6 $t = \dfrac{(12.6 - 13.7) - 0}{0.5} = -2.2$, with corresponding p-value .029 (from Table A.8)

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

43.

a. Although there is a "jump" in the middle of the Normal Probability plot, the data follow a reasonably straight path, so there is no strong reason for doubting the normality of the population of differences.

b. A 95% lower confidence bound for the population mean difference is:

$$\bar{d} - t_{.05,14}\left(\frac{s_d}{\sqrt{n}}\right) = -38.60 - (1.761)\left(\frac{23.18}{\sqrt{15}}\right) = -38.60 - 10.54 = -49.14.$$

Therefore, with a confidence level of 95%, the population mean difference is above (–49.14).

c. A 95% upper confidence bound for the corresponding population mean difference is $38.60 + 10.54 = 49.14$

44. We need to check the differences to see if the assumption of normality is plausible. A probability chart will validate our use of the t distribution. A 95% confidence interval:

$$\bar{d} + t_{.05,15}\left(\frac{s_d}{\sqrt{n}}\right) = 2635.63 + (1.753)\left(\frac{508.645}{\sqrt{16}}\right) = 2635.63 + 222.91$$

$$\Rightarrow (\infty, 2858.54)$$

45. The differences (white – black) are –7.62, -8.00, -9.09, -6.06, -1.39, -16.07, -8.40, -8.89, and –2.88, from which $\bar{d} = -7.600$, and $s_d = 4.178$. The confidence level is not specified in the problem description; for 95% confidence, $t_{.025,8} = 2.306$, and the C.I. is

$$-7.600 \pm (2.306)\left(\frac{4.178}{\sqrt{9}}\right) = -7.600 \pm 3.211 = (-10.811, -4.389).$$

46. 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), while $s_1 = s_2 = 8.96$, so $s_p = 8.96$ and t = .16.

Section 9.4

47. H_o will be rejected if $z \le -z_{.01} = -2.33$. With $\hat{p}_1 = .150$, and $\hat{p}_2 = .300$,

$$\hat{p} = \frac{30+80}{200+600} = \frac{210}{800} = .263, \text{ and } \hat{q} = .737.$$ The calculated test statistic is

$$z = \frac{.150-.300}{\sqrt{(.263)(.737)\left(\frac{1}{200}+\frac{1}{600}\right)}} = \frac{-.150}{.0359} = -4.18.$$ Because $-4.18 \le -2.33$, H_o is rejected; the proportion of those who repeat after inducement appears lower than those who repeat after no inducement.

48.

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

$$\hat{p} = \frac{63+75}{300+180} = .2875,\ z = \frac{.2100-.4167}{\sqrt{(.2875)(.7125)\left(\frac{1}{300}+\frac{1}{180}\right)}} = \frac{-.2067}{.0427} = -4.84.$$

Since $-4.84 \le -1.96$, H_o is rejected.

b. $\bar{p} = .275$ and $s = .0432$, so power =

$$1-\left[\Phi\left(\frac{[(1.96)(.0421)+.2]}{.0432}\right)-\Phi\left(\frac{[-(1.96)(.0421)+.2]}{.0432}\right)\right] =$$
$$1-[\Phi(6.54)-\Phi(2.72)] = .9967.$$

49.

1 Parameter of interest: $p_1 - p_2$ = true difference in proportions of those responding to two different survey covers. Let p_1 = Plain, p_2 = Picture.

2 $H_0: p_1 - p_2 = 0$

3 $H_a: p_1 - p_2 < 0$

4 $$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{m}+\frac{1}{n}\right)}}$$

5 Reject H_o if p-value < .10

6 $$z = \frac{\frac{104}{207}-\frac{109}{213}}{\sqrt{\left(\frac{213}{420}\right)\left(\frac{207}{420}\right)\left(\frac{1}{207}+\frac{1}{213}\right)}} = -.1910;$$ p-value = .4247

7 Fail to Reject H_o. The data does not indicate that plain cover surveys have a lower response rate.

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

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

51.

a. $H_0 : p_1 = p_2$ will be rejected in favor of $H_a : p_1 \neq p_2$ if either $z \geq 1.645$ or $z \leq -1.645$. With $\hat{p}_1 = .193$, and $\hat{p}_2 = .182$, $\hat{p} = .188$, $z = \dfrac{.011}{.00742} = 1.48$. Since 1.48 is not ≥ 1.645, H_o is not rejected and we conclude that no difference exists.

b. Using formula (9.7) with $p_1 = .2$, $p_2 = .18$, $a = .1$, $b = .1$, and $z_{a/2} = 1.645$,

$$n = \frac{\left(1.645\sqrt{.5(.38)(1.62)} + 1.28\sqrt{.16 + .1476}\right)^2}{.0004} = 6582$$

52. Let p_1 = true proportion of irradiated bulbs that are marketable; p_2 = true proportion of untreated bulbs that are marketable; The hypotheses are $H_0 : p_1 - p_2 = 0$ vs. $H_0 : p_1 - p_2 > 0$. The test statistic is $z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{m} + \frac{1}{n}\right)}}$. With $\hat{p}_1 = \dfrac{153}{180} = .850$, and $\hat{p}_2 = \dfrac{119}{180} = .661$, $\hat{p} = \dfrac{272}{360} = .756$, $z = \dfrac{.850 - .661}{\sqrt{(.756)(.244)\left(\frac{1}{180} + \frac{1}{180}\right)}} = \dfrac{.189}{.045} = 4.2$.

The p-value = $1 - \Phi(4.2) \approx 0$, so reject H_o at any reasonable level. Radiation appears to be beneficial.

53.

a. A 95% large sample confidence interval formula for $\ln(q)$ is

$\ln(\hat{q}) \pm z_{a/2}\sqrt{\dfrac{m-x}{mx} + \dfrac{n-y}{ny}}$. Taking the antilogs of the upper and lower bounds gives the confidence interval for q itself.

b. $\hat{q} = \dfrac{\frac{189}{11{,}034}}{\frac{104}{11{,}037}} = 1.818$, $\ln(\hat{q}) = .598$, and the standard deviation is

$\sqrt{\dfrac{10{,}845}{(11{,}034)(189)} + \dfrac{10{,}933}{(11{,}037)(104)}} = .1213$, so the CI for $\ln(q)$ is

$.598 \pm 1.96(.1213) = (.360, .836)$. Then taking the antilogs of the two bounds gives the CI for q to be $(1.43, 2.31)$.

54.

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 $(p_1 + p_3) – (p_1 + p_2)$ is $\dfrac{(X_1 + X_3) - (X_1 + X_2)}{n} = \dfrac{X_3 - X_2}{n}$.

c. When H_o is true, $p_2 = p_3$, so $Var\left(\dfrac{X_3 - X_2}{n}\right) = \dfrac{p_2 + p_3}{n}$, which is estimated by $\dfrac{\hat{p}_2 + \hat{p}_3}{n}$. The Z statistic is then $\dfrac{\frac{X_3 - X_2}{n}}{\sqrt{\frac{\hat{p}_2 + \hat{p}_3}{n}}} = \dfrac{X_3 - X_2}{\sqrt{X_2 + X_3}}$.

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

55. $\hat{p}_1 = \dfrac{15 + 7}{40} = .550$, $\hat{p}_2 = \dfrac{29}{42} = .690$, and the 95% C.I. is $(.550 - .690) \pm 1.96(.106) = -.14 \pm .21 = (-.35, .07)$.

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

Section 9.5

57.

a. From Table A.9, 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 $F_{.99,10,5} = \dfrac{1}{5.64} = .177$,
$P(.177 \le F \le 4.74) = P(F \le 4.74) - P(F \le .177) = .95 - .01 = .94$.

58.

a. Since the given f value of 4.75 falls between $F_{.05,5,10} = 3.33$ and $F_{.01,5,10} = 5.64$, we can say that the upper-tailed p-value is between .01 and .05.

b. Since the given f of 2.00 is less than $F_{.10,5,10} = 2.52$, the p-value > .10.

c. The two tailed p-value = $2P(F \ge 5.64) = 2(.01) = .02$.

d. For a lower tailed test, we must first use formula 9.9 to find the critical values:
$F_{.90,5,10} = \dfrac{1}{F_{.10,10,5}} = .3030$, $F_{.95,5,10} = \dfrac{1}{F_{.05,10,5}} = .2110$,
$F_{.99,5,10} = \dfrac{1}{F_{.01,10,5}} = .0995$. Since .0995 < f = .200 < .2110, .01 < p-value < .05 (but obviously closer to .05).

e. There is no column for numerator d.f. of 35 in Table A.9, however looking at both df = 30 and df = 40 columns, we see that for denominator df = 20, our f value is between $F_{.01}$ and $F_{.001}$. So we can say .001< p-value < .01.

59. We test $H_0 : s_1^2 = s_2^2$ vs. $H_a : s_1^2 \neq s_2^2$. The calculated test statistic is $f = \frac{(2.75)^2}{(4.44)^2} = .384$. With numerator d.f. = m – 1 = 10 – 1 = 9, and denominator d.f. = n – 1 = 5 – 1 = 4, we reject H_0 if $f \geq F_{.05,9,4} = 6.00$ or $f \leq F_{.95,9,4} = 1/F_{.05,4,9} = 1/3.63 = .275$. Since .384 is in neither rejection region, we do not reject H_0 and conclude that there is no significant difference between the two standard deviations.

60. With $s_1 =$ true standard deviation for not-fused specimens and $s_2 =$ true standard deviation for fused specimens, we test $H_0 : s_1 = s_2$ vs. $H_a : s_1 > s_2$. The calculated test statistic is $f = \frac{(277.3)^2}{(205.9)^2} = 1.814$. With numerator d.f. = m – 1 = 10 – 1 = 9, and denominator d.f. = 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_o. There is not sufficient evidence that the standard deviation of the strength distribution for fused specimens is smaller than that of not-fused specimens.

61. Let $s_1^2 =$ variance in weight gain for low-dose treatment, and $s_2^2 =$ variance in weight gain for control condition. We wish to test $H_0 : s_1^2 = s_2^2$ vs. $H_a : s_1^2 > s_2^2$. The test statistic is $f = \frac{s_1^2}{s_2^2}$, and we reject H_o at level .05 if $f > F_{.05,19,22} \approx 2.08$.

$f = \frac{(54)^2}{(32)^2} = 2.85 \geq 20.8$, so reject H_o at level .05. The data does suggest that there is more variability in the low-dose weight gains.

62. $H_0 : s_1 = s_2$ will be rejected in favor of $H_a : s_1 \neq s_2$ if either $f \leq F_{.975,47,44} \approx .56$ or if $f \geq F_{.025,47,44} \approx 1.8$. Because $f = 1.22$, H_o is not rejected. The data does not suggest a difference in the two variances.

63. $P\left(F_{1-a/2,m-1,n-1} \le \dfrac{S_1^2/s_1^2}{S_2^2/s_2^2} \le F_{a/2,m-1,n-1}\right) = 1-a$. The set of inequalities inside the parentheses is clearly equivalent to $\dfrac{S_2^2 F_{1-a/2,m-1,n-1}}{S_1^2} \le \dfrac{s_2^2}{s_1^2} \le \dfrac{S_2^2 F_{a/2,m-1,n-1}}{S_1^2}$. Substituting the sample values s_1^2 and s_2^2 yields the confidence interval for $\dfrac{s_2^2}{s_1^2}$, and taking the square root of each endpoint yields the confidence interval for $\dfrac{s_2}{s_1}$. m = n = 4, so we need $F_{.05,3,3} = 9.28$ and $F_{.95,3,3} = \dfrac{1}{9.28} = .108$. Then with s_1 = .160 and s_2 = .074, the C. I. for $\dfrac{s_2^2}{s_1^2}$ is (.023, 1.99), and for $\dfrac{s_2}{s_1}$ is (.15, 1.41).

64. A 95% upper bound for $\dfrac{s_2}{s_1}$ is $\sqrt{\dfrac{s_2^2 F_{.05,9,9}}{s_1^2}} = \sqrt{\dfrac{(3.59)^2(3.18)}{(.79)^2}} = 8.10$. We are confident that the ratio of the standard deviation of triacetate porosity distribution to that of the cotton porosity distribution is at most 8.10.

Supplementary Exercises

65. We test $H_0 : m_1 - m_2 = 0$ vs. $H_a : m_1 - m_2 \ne 0$. The test statistic is

$$t = \frac{(\bar{x}-\bar{y})-(\Delta)}{\sqrt{\frac{s_1^2}{m}+\frac{s_2^2}{n}}} = \frac{807-757}{\sqrt{\frac{27^2}{10}+\frac{41^2}{10}}} = \frac{50}{\sqrt{241}} = \frac{50}{15.524} = 3.22$$

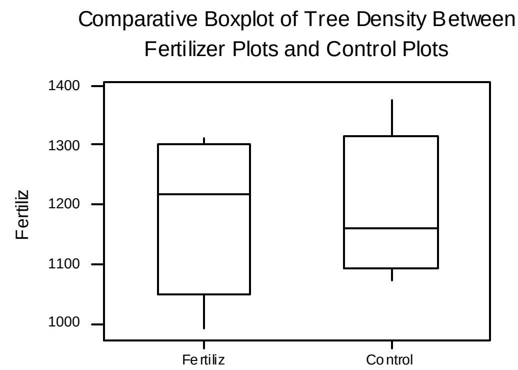
The approximate d.f. is

$$n = \frac{(241)^2}{\frac{(72.9)^2}{9}+\frac{(168.1)^2}{9}} = 15.6$$

, which we round down to 15. The p-value for a two-tailed test is approximately 2P(t > 3.22) = 2(.003) = .006. This small of a p-value gives strong support for the alternative hypothesis. The data indicates a significant difference.

66.

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 $H_0 : m_1 - m_2 = 0$ vs. $H_a : m_1 - m_2 \neq 0$ yields a t value of -.20, and a two-tailed p-value of .85. (d.f. = 13). We would fail to reject H_o; 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.

67. Let p_1 = true proportion of returned questionnaires that included no incentive; p_2 = true proportion of returned questionnaires that included an incentive. The hypotheses are $H_0 : p_1 - p_2 = 0$ vs. $H_0 : p_1 - p_2 < 0$. The test statistic is $z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{m} + \frac{1}{n}\right)}}$.

$\hat{p}_1 = \dfrac{75}{110} = .682$, and $\hat{p}_2 = \dfrac{66}{98} = .673$. At this point we notice that since $\hat{p}_1 > \hat{p}_2$, the numerator of the z statistic will be > 0, and since we have a lower tailed test, the p-value will be > .5. We fail to reject H_o. This data does not suggest that including an incentive increases the likelihood of a response.

68. Summary quantities are m = 24, $\bar{x} = 103.66$, $s_1 = 3.74$, n = 11, $\bar{y} = 101.11$, $s_2 = 3.60$. We use the pooled t interval based on 24 + 11 – 2 = 33 d.f.; 95% confidence requires $t_{.025,33} = 2.03$. With $s_p^2 = 13.68$ and $s_p = 3.70$, the confidence interval is $2.55 \pm (2.03)(3.70)\sqrt{\frac{1}{24}+\frac{1}{11}} = 2.55 \pm 2.73 = (-.18, 5.28)$. We are confident that the difference between true average dry densities for the two sampling methods is between -.18 and 5.28. Because the interval contains 0, we cannot say that there is a significant difference between them.

69. The center of any confidence interval for $m_1 - m_2$ 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 $609.3 \pm (1.645)(552.35) = 609.3 \pm 908.6$ $= (-299.3, 1517.9)$.

70.

a. A 95% lower confidence bound for the true average strength of joints with a side coating is $\bar{x} - t_{.025,9}\left(\dfrac{s}{\sqrt{n}}\right) = 63.23 - (1.833)\left(\dfrac{5.96}{\sqrt{10}}\right) = 63.23 - 3.45 = 59.78$. That is, with a confidence level of 95%, the average strength of joints with a side coating is at least 59.78 (Note: this bound is valid only if the distribution of joint strength is normal.)

b. A 95% lower prediction bound for the strength of a single joint with a side coating is $\bar{x} - t_{.025,9}\left(s\sqrt{1+\frac{1}{n}}\right) = 63.23 - (1.833)\left(5.96\sqrt{1+\frac{1}{10}}\right) = 63.23 - 11.46 = 51.77$. That is, with a confidence level of 95%, the strength of a single joint with a side coating would be at least 51.77.

c. For a confidence level of 95%, a two-sided tolerance interval for capturing at least 95% of the strength values of joints with side coating is $\bar{x} \pm$ (tolerance critical value)s. The tolerance critical value is obtained from Table A.6 with 95% confidence, k = 95%, and n = 10. Thus, the interval is $63.23 \pm (3.379)(5.96) = 63.23 \pm 20.14 = (43.09, 83.37)$. That is, we can be highly confident that at least 95% of all joints with side coatings have strength values between 43.09 and 83.37.

d. A 95% confidence interval for the difference between the true average strengths for the two types of joints is $(80.95-63.23)\pm t_{.025,\nu}\sqrt{\frac{(9.59)^2}{10}+\frac{(5.96)^2}{10}}$. The approximate degrees of freedom is $\nu=\frac{\left(\frac{91.9681}{10}+\frac{35.5216}{10}\right)^2}{\frac{\left(\frac{91.9681}{10}\right)^2}{9}+\frac{\left(\frac{35.5216}{10}\right)^2}{9}}=15.05$, so we use 15 d.f., and $t_{.025,15}=2.131$. The interval is , then,

$17.72\pm(2.131)(3.57)=17.72\pm 7.61=(10.11,25.33)$. With 95% confidence, we can say that the true average strength for joints without side coating exceeds that of joints with side coating by between 10.11 and 25.33 lb-in./in.

71. 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 $(3975.0-2795.0)\pm 2.58\sqrt{\frac{245.1^2}{40}+\frac{293.7^2}{40}}$

$(1180.0)\pm 1560.5\approx(1024,1336)$. 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.

72. This exercise calls for a paired analysis. First compute the difference between the amount of cone penetration for commutator and pinion bearings for each of the 17 motors. These 17 differences are summarized as follows: n = 17, $\bar{d}=-4.18$, $s_d=35.85$, where d = (commutator value – pinion value). Then $t_{.025,16}=2.120$, and the 95% confidence interval for the population mean difference between penetration for the commutator armature bearing and penetration for the pinion bearing is:

$-4.18\pm(2.120)\left(\frac{35.85}{\sqrt{17}}\right)=-4.18\pm 18.43=(-22.61,14.25)$. We would have to say that the population mean difference has not been precisely estimated. The bound on the error of estimation is quite large. In addition, the confidence interval spans zero. Because of this, we have insufficient evidence to claim that the population mean penetration differs for the two types of bearings.

73. Since we can assume that the distributions from which the samples were taken are normal, we use the two-sample t test. Let m_1 denote the true mean headability rating for aluminum killed steel specimens and m_2 denote the true mean headability rating for silicon killed steel. Then the hypotheses are $H_0 : m_1 - m_2 = 0$ vs. $H_a : m_1 - m_2 \neq 0$. The test statistic is $t = \frac{-.66}{\sqrt{.03888 + .047203}} = \frac{-.66}{\sqrt{.086083}} = -2.25$. The approximate degrees of freedom

$$n = \frac{(.086083)^2}{\frac{(.03888)^2}{29} + \frac{(.047203)^2}{29}} = 57.5$$

, so we use 57. The two-tailed p-value $\approx 2(.014) = .028$, which is less than the specified significance level, so we would reject H_o. The data supports the article's authors' claim.

74. Let m_1 denote the true average tear length for Brand A and let m_2 denote the true average tear length for Brand B. The relevant hypotheses are $H_0 : m_1 - m_2 = 0$ vs. $H_a : m_1 - m_2 > 0$. 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 d.f. is

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

, 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_o is rejected and we conclude that the average tear length for Brand A is larger than that of Brand B.

75.

a. The relevant hypotheses are $H_0 : m_1 - m_2 = 0$ vs. $H_a : m_1 - m_2 \neq 0$. Assuming both populations have normal distributions, the two-sample t test is appropriate. m = 11, $\bar{x} = 98.1$, s_1 = 14.2, n = 15, $\bar{y} = 129.2$, s_2 = 39.1. The test statistic is

$$t = \frac{-31.1}{\sqrt{18.3309 + 101.9207}} = \frac{-31.1}{\sqrt{120.252}} = -2.84$$

. The approximate degrees of freedom

$$n = \frac{(120.252)^2}{\frac{(18.3309)^2}{10} + \frac{(101.9207)^2}{14}} = 18.64$$

, so we use 18. From Table A.7, the two-tailed p-value $\approx 2(.006) = .012$. No, obviously, the results are different.

b. For the hypotheses $H_0: m_1 - m_2 = -25$ vs. $H_a: m_1 - m_2 < -25$, the test statistic changes to $t = \dfrac{-31.1 - (-25)}{\sqrt{120.252}} = -.556$. With degrees of freedom 18, the p-value $\approx P(t < -.6) = .278$. Since the p-value is greater than any sensible choice of a, we fail to reject H_o. There is insufficient evidence that the true average strength for males exceeds that for females by more than 25N.

76.

a. The relevant hypotheses are $H_0: m_1^* - m_2^* = 0$ (which is equivalent to saying $m_1 - m_2 = 0$) versus $H_a: m_1^* - m_2^* \neq 0$ (which is the same as saying $m_1 - m_2 \neq 0$). The pooled t test is based on d.f. = m + n – 2 = 8 + 9 – 2 = 15. The pooled variance is $s_p^2 = \left(\dfrac{m-1}{m+n-2}\right)s_1^2 + \left(\dfrac{n-1}{m+n-2}\right)s_2^2$

$\left(\dfrac{8-1}{8+9-2}\right)(4.9)^2 + \left(\dfrac{9-1}{8+9-2}\right)(4.6)^2 = 22.49$, so $s_p = 4.742$. The test statistic is $t = \dfrac{\bar{x}^* - \bar{y}^*}{s_p\sqrt{\frac{1}{m}+\frac{1}{n}}} = \dfrac{18.0 - 11.0}{4.742\sqrt{\frac{1}{8}+\frac{1}{9}}} = 3.04 \approx 3.0$. From Table A.7, the p-value associated with t = 3.0 is 2P(t > 3.0) = 2(.004) = .008. At significance level .05, H_o is rejected and we conclude that there is a difference between m_1^* and m_2^*, which is equivalent to saying that there is a difference between m_1 and m_2.

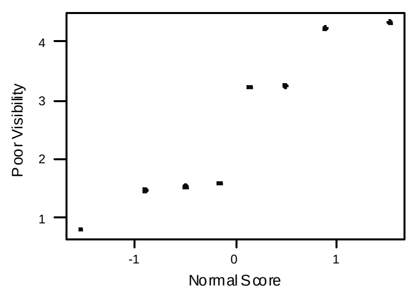
b. No. The mean of a lognormal distribution is $m = e^{m^* + (s^*)^2/2}$, where m^* and s^* are the parameters of the lognormal distribution (i.e., the mean and standard deviation of ln(x)). So when $s_1^* = s_2^*$, then $m_1^* = m_2^*$ would imply that $m_1 = m_2$. However, when $s_1^* \neq s_2^*$, then even if $m_1^* = m_2^*$, the two means m_1 and m_2 (given by the formula above) would not be equal.

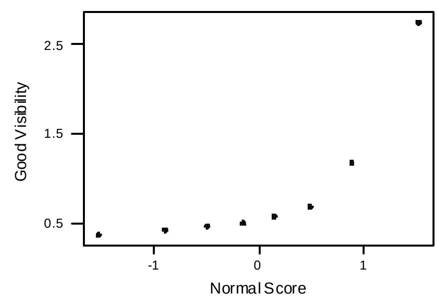
77. This is paired data, so the paired t test is employed. The relevant hypotheses are $H_0: m_d = 0$ vs. $H_a: m_d < 0$, where m_d denotes the difference between the population average control strength minus the population average heated strength. The observed differences (control – heated) are: -.06, .01, -.02, 0, and -.05. The sample mean and standard deviation of the differences are $\bar{d} = -.024$ and $s_d = .0305$. The test statistic is

$t = \dfrac{-.024}{.0305/\sqrt{5}} = -1.76 \approx -1.8$. From Table A.7, with d.f. = 5 – 1 = 4, the lower tailed p-value associated with t = -1.8 is P(t < -1.8) = P(t > 1.8) = .073. At significance level .05, H_o should not be rejected. Therefore, this data does not show that the heated average strength exceeds the average strength for the control population.

78. Let m_1 denote the true average ratio for young men and m_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 $H_0 : m_1 - m_2 = 0$ vs. $H_a : m_1 - m_2 > 0$. 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 d.f. = 20 and the p-value is P(t > 7.5) ≈ 0. Since the p-value is $< a = .05$, we reject H$_o$. We have sufficient evidence to claim that the true average ratio for young men exceeds that for elderly men.

79.





A normal probability plot indicates the data for good visibility does not follow a normal distribution, thus a t-test is not appropriate for this small a sample size.

80. The relevant hypotheses would be $m_M = m_F$ versus $m_M \neq m_F$ for both the distress and delight indices. The reported p-value for the test of mean differences on the distress index was less than 0.001. This indicates a statistically significant difference in the mean scores, with the mean score for women being higher. The reported p-value for the test of mean differences on the delight index was > 0.05. This indicates a lack of statistical significance in the difference of delight index scores for men and women.

81. We wish to test H$_0$: $m_1 = m_2$ versus H$_a$: $m_1 \neq m_2$

Unpooled:

With H$_o$: $m_1 - m_2 = 0$ vs. H$_a$: $m_1 - m_2 \neq 0$, we will reject H$_o$ if $p - value < a$.

$$n = \frac{\left(\frac{.79^2}{14} + \frac{1.52^2}{12}\right)^2}{\frac{\left(\frac{.79^2}{14}\right)^2}{13} + \frac{\left(\frac{1.52^2}{12}\right)^2}{11}} = 15.95 \approx 16$$, and the test statistic

$$t = \frac{8.48 - 9.36}{\sqrt{\frac{.79^2}{14} + \frac{1.52^2}{12}}} = \frac{-.96}{.4869} = -1.97$$ leads to a p-value of 2[P(t > 1.97)] $\approx 2(.031) \approx .062$

Pooled:

The degrees of freedom $n = m = n - 2 = 14 + 12 - 2 = 24$ and the pooled variance is $\left(\frac{13}{24}\right)(.79)^2 + \left(\frac{11}{24}\right)(1.52)^2 = 1.3970$, so $s_p = 1.181$. The test statistic is

$$t = \frac{-.96}{1.181\sqrt{\frac{1}{14} + \frac{1}{12}}} = \frac{-.96}{.465} \approx -2.1$$. The p-value = 2[P(t_{24} > 2.1)] = 2(.023) = .046.

With the pooled method, there are more degrees of freedom, and the p-value is smaller than with the unpooled method.

82. 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$_a$: μ_d ? 0. Test statistic $t = \frac{1.757}{1.197/\sqrt{7}} = 3.88$ with df = n – 1 = 6 leads to a p-value of 2[P(t > 3.88) ˜ .004. 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.

83.

a. With n denoting the second sample size, the first is m = 3n. We then wish $20 = 2(2.58)\sqrt{\frac{900}{3n} + \frac{400}{n}}$, which yields n = 47, m = 141.

b. We wish to find the n which minimizes $2(z_{a/2})\sqrt{\frac{900}{400-n} + \frac{400}{n}}$, or equivalently, the n which minimizes $\frac{900}{400-n} + \frac{400}{n}$. Taking the derivative with respect to n and equating to 0 yields $900(400-n)^{-2} - 400n^{-2} = 0$, whence $9n^2 = 4(400-n)^2$, or $5n^2 + 3200n - 640{,}000 = 0$. This yields n = 160, m = 400 – n = 240.

84. Let p_1 = true survival rate at $11^{o}C$; p_2 = true survival rate at $30^{o}C$; The hypotheses are $H_0 : p_1 - p_2 = 0$ vs. $H_a : p_1 - p_2 \neq 0$. The test statistic is $z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{m}+\frac{1}{n}\right)}}$. With

$\hat{p}_1 = \frac{73}{91} = .802$, and $\hat{p}_2 = \frac{102}{110} = .927$, $\hat{p} = \frac{175}{201} = .871$, $\hat{q} = .129$.

$z = \frac{.802 - .927}{\sqrt{(.871)(.129)\left(\frac{1}{91}+\frac{1}{110}\right)}} = \frac{-.125}{.0320} = -3.91$. The p-value =

$\Phi(-3.91) < \Phi(-3.49) = .0003$, so reject H_o at any reasonable level. The two survival rates appear to differ.

85.

a. We test $H_0 : m_1 - m_2 = 0$ vs. $H_a : m_1 - m_2 \neq 0$. Assuming both populations have normal distributions, the two-sample t test is appropriate. The approximate degrees of freedom $n = \frac{(.042721)^2}{\frac{(.0325125)^2}{7} + \frac{(.0102083)^2}{11}} = 11.4$, so we use df = 11.

$t_{.0005,11} = 4.437$, so we reject H_o if $t \geq 4.437$ or $t \leq -4.437$ The test statistic is $t = \frac{.68}{\sqrt{.042721}} \approx 3.3$, which is not ≥ 4.437, so we cannot reject H_o. At significance level .001, the data does not indicate a difference in true average insulin-binding capacity due to the dosage level.

b. P-value = 2P(t > 3.3) = 2 (.004) = .008 which is > .001.

86. $\hat{s}^2 = \frac{\left[(n_1-1)S_1^2 + (n_2-1)S_2^2 + (n_3-1)S_3^2 + (n_4-1)S_4^2\right]}{n_1+n_2+n_3+n_4-4}$

$E(\hat{s}^2) = \frac{\left[(n_1-1)s_1^2 + (n_2-1)s_2^2 + (n_3-1)s_3^2 + (n_4-1)s_4^2\right]}{n_1+n_2+n_3+n_4-4} = s^2$. The estimate for the given data is $= \frac{[15(.4096)+17(.6561)+7(.2601)+11(.1225)]}{50} = .409$

87. $\Delta_0 = 0$, $s_1 = s_2 = 10$, d = 1, $s = \sqrt{\frac{200}{n}} = \frac{14.142}{\sqrt{n}}$, so $b = \Phi\left(1.645 - \frac{\sqrt{n}}{14.142}\right)$, giving $b =$ .9015, .8264, .0294, and .0000 for n = 25, 100, 2500, and 10,000 respectively. If the m_i's referred to true average IQ's resulting from two different conditions, $m_1 - m_2 = 1$ would have little practical significance, yet very large sample sizes would yield statistical significance in this situation.

88. $H_0: m_1 - m_2 = 0$ is tested against $H_a: m_1 - m_2 \neq 0$ using the two-sample t test, rejecting H_o at level .05 if either $t \geq t_{.025,15} = 2.131$ or if $t \leq -2.131$. With $\bar{x} = 11.20$, $s_1 = 2.68$, $\bar{y} = 9.79$, $s_2 = 3.21$, and m = n = 8, s_p = 2.96, and t = .95, so H_o is not rejected. In the situation described, the effect of carpeting would be mixed up with any effects due to the different types of hospitals, so no separate assessment could be made. The experiment should have been designed so that a separate assessment could be obtained (e.g., a randomized block design).

89. $H_0: p_1 = p_2$ will be rejected at level a in favor of $H_a: p_1 > p_2$ if either $z \geq z_{.05} = 1.645$. With $\hat{p}_1 = \frac{250}{2500} = .10$, $\hat{p}_2 = \frac{167}{2500} = .0668$, and $\hat{p} = .0834$, $z = \frac{.0332}{.0079} = 4.2$, so H_o is rejected. It appears that a response is more likely for a white name than for a black name.

90. The computed value of Z is $z = \frac{34 - 46}{\sqrt{34 + 46}} = -1.34$. A lower tailed test would be appropriate, so the p-value $= \Phi(-1.34) = .0901 > .05$, so we would not judge the drug to be effective.

91.

a. Let m_1 and m_2 denote the true average weights for operations 1 and 2, respectively. The relevant hypotheses are $H_0 : m_1 - m_2 = 0$ vs. $H_a : m_1 - m_2 \neq 0$. The value of the test statistic is

$$t = \frac{(1402.24 - 1419.63)}{\sqrt{\frac{(10.97)^2}{30} + \frac{(9.96)^2}{30}}} = \frac{-17.39}{\sqrt{4.011363 + 3.30672}} = \frac{-17.39}{\sqrt{7.318083}} = -6.43.$$

The d.f. $n = \dfrac{(7.318083)^2}{\frac{(4.011363)^2}{29} + \frac{(3.30672)^2}{29}} = 57.5$, so use df = 57. $t_{.025,57} \approx 2.000$, so we can reject H_o at level .05. The data indicates that there is a significant difference between the true mean weights of the packages for the two operations.

b. $H_0 : m_1 = 1400$ will be tested against $H_a : m_1 > 1400$ using a one-sample t test with test statistic $t = \dfrac{\bar{x} - 1400}{s_1 / \sqrt{m}}$. With degrees of freedom = 29, we reject H_o if $t > t_{.05,29} = 1.699$. The test statistic value is $t = \dfrac{1402.24 - 1400}{10.97/\sqrt{30}} = \dfrac{2.24}{2.00} = 1.1$.

Because 1.1 < 1.699, H_o is not rejected. True average weight does not appear to exceed 1400.

92. $Var(\bar{X} - \bar{Y}) = \dfrac{l_1}{m} + \dfrac{l_2}{n}$ and $\hat{l}_1 = \bar{X}$, $\hat{l}_2 = \bar{Y}$, $\hat{l} = \dfrac{m\bar{X} + n\bar{Y}}{m + n}$, giving $Z = \dfrac{\bar{X} - \bar{Y}}{\sqrt{\frac{\hat{l}}{m} + \frac{\hat{l}}{n}}}$. With $\bar{x} = 1.616$ and $\bar{y} = 2.557$, z = -5.3 and p-value = $2(\Phi(-5.3)) < .0006$, so we would certainly reject $H_0 : l_1 = l_2$ in favor of $H_a : l_1 \neq l_2$.

93. $\hat{l}_1 = \bar{x} = 1.62$, $\hat{l}_2 = \bar{y} = 2.56$, $\sqrt{\dfrac{\hat{l}_1}{m} + \dfrac{\hat{l}_2}{n}} = 1.77$, and the confidence interval is

$$-.94 \pm (1.96)(1.77) = -.94 \pm .35 = (-1.29, -.59)$$