Chapter 13

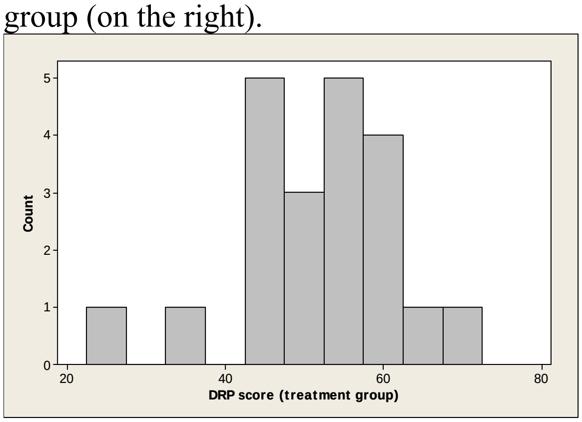
13.1 (a) Counts will be obtained from the samples so this is a problem about comparing proportions. (b) This is an observational study comparing random samples selected from two independent populations.

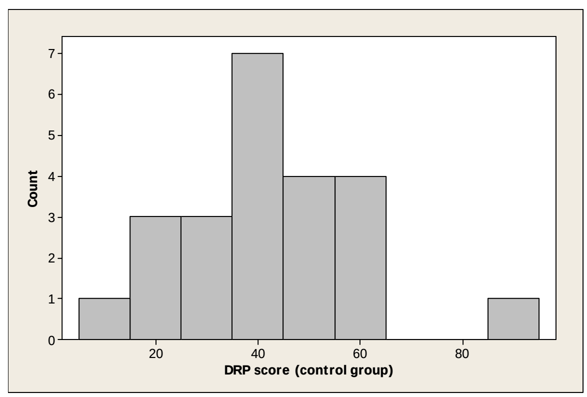
13.2 (a) Scores will be obtained from the samples so this is a problem about comparing means (average scores). (b) This is an experiment because the researchers an imposing a "treatment" and measuring a response variable. Since these are volunteers we will not be able to generalize the results to all gamers.

13.3 (a) Two samples. The two segments are used by two independent groups of children. (b) Paired data. The two segments are both used by each child.

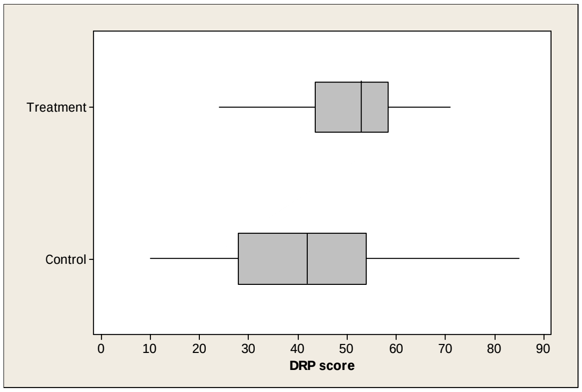
13.4 (a) Single sample. The sample mean will be compared with the known concentration. (b) Two samples. The mean concentration in 10 beakers with the new method will be compared to the mean concentration in 10 different beakers with the old method

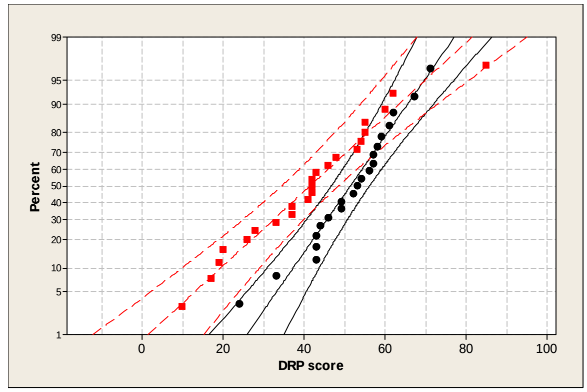
13.5 (a) $H_0: \mu_T = \mu_C$ versus $H_a: \mu_T > \mu_C$, where μ_T and μ_C are the mean improvement of reading ability of the treatment and control group respectively. (b) The treatment group is slightly left-skewed with a greater mean and smaller standard deviation ($\bar{x}$=51.48, s= 11.01) than the control group ($\bar{x}$=41.52, s= 17.15). The histograms below show no serious departures from Normality for the treatment group (on the left) and one unusually large score for the control group (on the right).





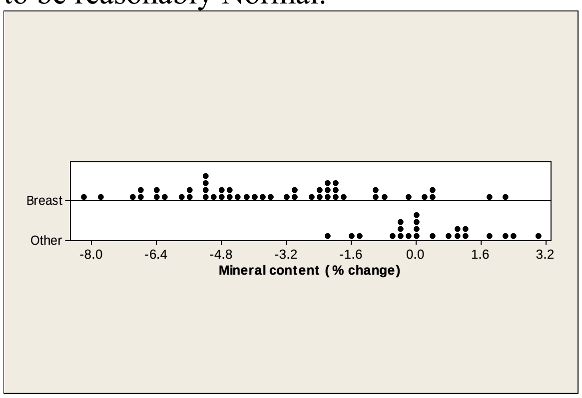
The boxplot (on the left below) also shows that the median DRP score is higher for the treatment group and the IQR is higher for the control group. Notice that the unusually high score is not identified as an outlier by Minitab. The combined Normal probability plot (on the right below) shows an overall liner trend for both sets of scores, so the Normal condition is satisfied for both groups.

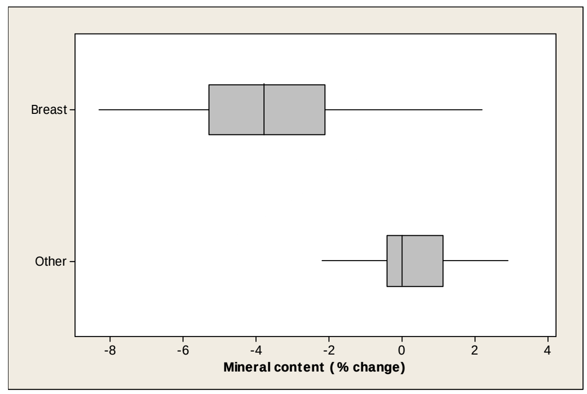




(c) Randomization was not possible, because existing classes were used. The researcher could not randomly assign the students to the two groups without disrupting classes.

13.6 (a) The two populations are breast-feeding women and other women. We want to test $H_0: \mu_B = \mu_C$ versus $H_a: \mu_B < \mu_C$, where μ_B and μ_C are the mean percent change in mineral content of the spines over three months for breast-feeding and other mothers, respectively.
(b) Dotplots (on the left) and boxplots (on the right) are shown below. Both distributions appear to be reasonably Normal.





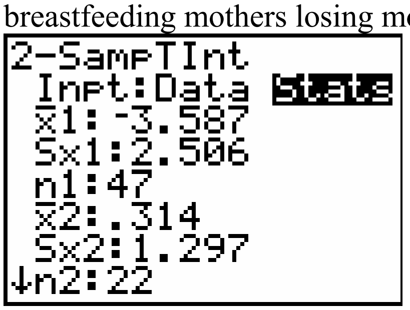
Breast-feeding mothers have a lower mean mineral content ($\bar{x} = -3.587$, s= 2.506) with more variability than other mothers ($\bar{x} = 0.314$, s= 1.297). (c) This is an observational study so we cannot make a cause and effect conclusion, but this effect is certainly worth investigating because there appears to be a difference in the two groups of mothers for some reason.

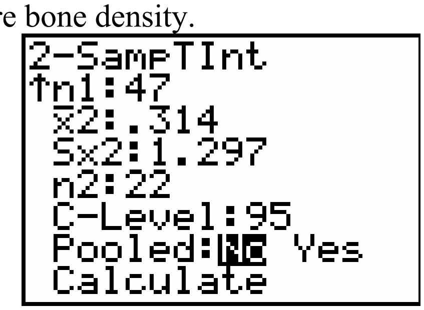
13.7 (a) The hypotheses should involve μ_1 and μ_2 (population means) rather than $\bar{x}_1$ and $\bar{x}_2$ (sample means). (b) The samples are not independent. We would need to compare the scores of the 10 boys to the scores for the 10 girls. (c) We need the P-value to be small (for example, less than 0.05) to reject H_0. A large P-value like this gives no reason to doubt H_0.

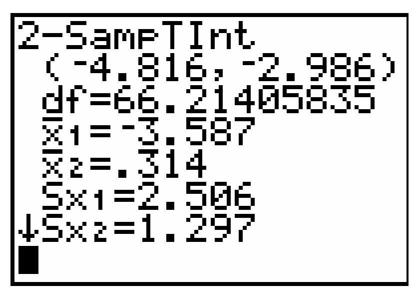
13.8 (a) Answers will vary. Examine random digits, if the digit is even then use Design A, otherwise use Design B. Once you use a design 30 days, stop and use the other design for the remaining days in the study. The first three digits are even, so the first three days for using Design A would be days 1, 2, and 3. (Note, if Design A is used when the digit is odd, then the first three days for using Design A are day 5, day 6, and day 8.) (b) Use a two-sided alternative ($H_0: \mu_A = \mu_B$ versus $H_a: \mu_A \neq \mu_B$), because we (presumably) have no prior suspicion that one design will be better than the other. (c) Both sample sizes are the same ($n_1 = n_2 = 30$), so the appropriate degrees of freedom would be df = 30 − 1 = 29. (d) Because 2.045 < t < 2.150, and the alternative is two-sided, Table C tells us that 0.04 < P-value < 0.05. (Software gives P = 0.0485.) We would reject H_0 and conclude that there is a difference in the mean daily sales for the two designs.

13.9 (a) We want to test $H_0: \mu_T = \mu_C$ versus $H_a: \mu_T > \mu_C$. The test statistic is $t = \dfrac{51.48 - 41.52}{\sqrt{11.01^2/21 + 17.15^2/23}} \doteq 2.311$, 0.01 < P-value < 0.02 with df = 20 (TI calculator gives P-value = 0.0132 with df = 37.86 and Minitab gives P-value = 0.013 with df=37). At the 5% significance level, it does not matter which method you use to obtain the P-value. The P-value (rounded to 0.013) is less than 0.05, so the data give good evidence that the new activities improve the mean DRP score. (b) A 95% confidence interval for $\mu_T - \mu_C$ is $(51.48 - 41.52) \pm 2.086\sqrt{11.01^2/21 + 17.15^2/23}$ = (0.97, 18.94) with df = 20; (1.233, 18.68) on TI calculator with df = 37.86; and (1.22637, 18.68254) using Minitab with df = 37. We estimate the mean improvement in reading ability using the new reading activities compared to not using them over an 8-week period to be between 1.23 and 18.68 points.

13.10 We want to test $H_0: \mu_B = \mu_C$ versus $H_a: \mu_B < \mu_C$. The test statistic is $t = \dfrac{-3.59 - 0.31}{\sqrt{2.51^2/47 + 1.30^2/22}} \doteq -8.51$, P-value < 0.0005 with df = 21 (the TI calculator and Minitab give P-values very close to 0). The small P-value is less than any reasonable significance level, say 1%, so the data give very strong evidence that nursing mothers on average lose more bone mineral than other mothers. (b) A 95% confidence interval for $\mu_B - \mu_O$ is $(-3.59 - 0.31) \pm 2.080\sqrt{2.51^2/47 + 1.30^2/22}$ = (−4.86, −2.95) with df = 21; (−4.816, −2.986) on TI calculator with df =66.21 (see the screen shots below); and (−4.81632, −2.98633) using Minitab with df = 66. We estimate the difference in the mean change in bone mineral for breastfeeding mothers when compared to other mothers to be between about 3% and 5%, with breastfeeding mothers losing more bone density.







13.11 (a) Because the sample sizes are so large, the t procedures are robust against non-Normality in the populations. (b) A 90% confidence interval for $\mu_M - \mu_F$ is $(1884.52 - 1360.39) \pm 1.660\sqrt{1368.37^2/675 + 1037.46^2/621}$ = (\$412.68, 635.58) using df = 100; (\$413.54, \$634.72) using df = 620; (\$413.62, 634.64) using df = 1249.21. We are 90% confident that the difference in mean summer earnings is between \$413.62 and \$634.64 higher for men. (c) The sample is not really random, but there is no reason to expect that the method used should introduce any bias. This is known as systematic sampling. (d) Students without employment were excluded, so the survey results can only (possibly) extend to employed undergraduates. Knowing the number of unreturned questionnaires would also be useful. These students are from one college, so it would be very helpful to know if this student body is representative of some larger group of students. It is very unlikely that you will be able to generalize these results to all undergraduates.

13.12 Answers will vary.

13.13 (a) We want to test $H_0: \mu_R = \mu_W$ versus $H_a: \mu_R > \mu_W$, where μ_R and μ_W are the mean percent change in polyphenols for men who drink red and white wine respectively. The test statistic is $t = \dfrac{5.5 - 0.23}{\sqrt{2.52^2/9 + 3.29^2/9}} \doteq 3.81$ with df = 8 and 0.0025 < P-value < 0.005. (b) The value of the test statistic is the same, but df = 14.97 and the P-value is 0.00085 (Minitab gives 0.001 with df = 14). The more complicated degrees of freedom give a smaller and less conservative P-value. (c) This study appears to have been a well-designed experiment, so it does provide evidence of causation.

13.14 (a) A 95% confidence interval for $\mu_R - \mu_W$ is $(5.5 - 0.23) \pm 2.306\sqrt{2.52^2/9 + 3.29^2/9} =$ (2.08%, 8.45%). (b) With df = 14.97, $t^* = 2.132$ and the confidence interval is 2.32% to 8.21%. (Minitab gives 2.304% to 8.229% with df = 14.) There is very little difference in the resulting confidence intervals.

13.15 (a) We want to test $H_0: \mu_S = \mu_N$ versus $H_a: \mu_S > \mu_N$, where μ_S and μ_N are the mean knee velocities for skilled and novice female competitive rowers, respectively. The test statistic is $t = 3.1583$ and the P-value = 0.0052. Note that the two-sided P-value is provided on the SAS output, so to get the appropriate P-value for the one-sided test use 0.0104/2 = 0.0052. Since 0.0052 < 0.01, we reject H_0 at the 1% level and conclude that the mean knee velocity is higher for skilled rowers. (b) Using df = 9.2, the critical value is $t^* = 1.8162$ and the resulting confidence interval for $\mu_S - \mu_N$ is (0.4982, 1.8475). With 90% confidence, we estimate that skilled female rowers have a mean angular knee velocity of between 0.498 and 1.847 units higher than that of novice female rowers. (c) Taking the conservative approach with Table C, df = 7 and the critical value is $t^* = 1.895$. Since 1.895 > 1.8162, the margin of error would be larger, so the confidence interval would be slightly wider.

13.16 (a) The missing t statistic is $t = \dfrac{70.37 - 68.45}{\sqrt{6.10035^2/10 + 9.03999^2/8}} \doteq 0.5143$. (b) We want to test $H_0: \mu_S = \mu_N$ versus $H_a: \mu_S \neq \mu_N$, where μ_S and μ_N are the mean weights of skilled and novice female competitive rowers, respectively. The test statistic is $t = 0.5143$ and the P-value = 0.6165. Since 0.6165 > 0.05, we cannot reject H_0 at the 5% level. There is no significant difference in the mean weights for skilled and novice rowers. (c) The more conservative approach would use df = 7. The t distribution with df = 7 has slightly heavier tails than the t distribution with df = 11.2, so the conservative P-value would be larger.

13.17 (a) Two-sample t test. (b) Paired t test. (c) Paired t test. (d) Two-sample t test. (e) Paired t test.

13.18 (a) The summary table is shown below. The only values not given directly are the standard deviations, which are found by computing $s = \sqrt{10}SEM$. (b) Use df = 9.

Group	Treatment	n	$\bar{x}$	s
1	IDX	10	116.0	17.71
2	Untreated	10	88.5	6.01

(c) This is a completely randomized design with one control group and one treatment group. The easiest way to carry out the randomization might be to number the hamsters (or their individual cages) from 1 to 20. Use the SRS applet and put 20 balls in the population hopper. Select 10 balls from the hopper. The 10 hamsters with these numbers will be injected with IDX. The other 10 hamsters will serve as the control group.

13.19 (a) Yes, the test statistic for testing $H_0: \mu_1 = \mu_2$ versus $H_a: \mu_1 > \mu_2$ is $t = \dfrac{116 - 88.5}{\sqrt{17.71^2/10 + 6.01^2/10}} \doteq 4.65$. With either df = 9 or df = 11.05, we have a significant result (P-value < 0.001 or P-value < 0.0005, respectively), so there is strong evidence that IDX prolongs life. (b) If using df = 9, the 95% confidence interval for $\mu_1 - \mu_2$ is $(116 - 88.5) \pm 2.262\sqrt{17.71^2/10 + 6.01^2/10}$ = (14.12, 40.88). With 95% confidence we estimate that IDX hamsters live, on average, between 14.12 and 40.88 days longer than untreated hamsters. If using df = 11.05, the interval is (14.49, 40.51).

13.20 (a) This is a two-sample t statistic, comparing two independent groups (supplemented and control). (b) Using the conservative df = 5, $t = -1.05$ would have a P-value between 0.30 and 0.40, which (as the report said) is not significant.

13.21 We want to test $H_0: \mu_C = \mu_S$ versus $H_a: \mu_C \neq \mu_S$. The test statistic is $t = \dfrac{4.0 - 11.3}{\sqrt{3.10934^2/6 + 3.92556^2/7}} \doteq -3.74$ and the P-value is between 0.01 and 0.02 (df = 5) or 0.0033 (df = 10.95), agreeing with the stated conclusion (a significant difference).

13.22 (a) These are paired t statistics: For each bird, the number of days behind the caterpillar peak was observed, and the t values were computed based on the pairwise differences between the first and second years. (b) For the control group, df = 5, and for the supplemented group, df = 6. (c) The control t is not significant (so the birds in that group did *not* "advance their laying date in the second year"), while the supplemented group t is significant with a one-sided P-value = 0.0195 (so those birds did change their laying date).

13.23 Answers will vary, but here is an example. The difference between average female (55.5) and male (57.9) self-concept scores was so small that it can be attributed to chance variation in the samples ($t = -0.83$, df = 62.8, P-value = 0.4110). In other words, based on this sample, we have no evidence that mean self-concept scores differ by gender.

13.24 (a) If the loggers had known that a study would be done, they might have (consciously or subconsciously) cut down fewer trees, in order to reduce the impact of logging. (b) Random assignment allows us to make a cause and effect conclusion. (c) We want to test $H_0: \mu_U = \mu_L$ versus $H_a: \mu_U > \mu_L$, where μ_U and μ_L are the mean number of species in unlogged and logged plots respectively. The test statistic is $t = \dfrac{17.5-13.67}{\sqrt{3.53^2/12+4.5^2/9}} \doteq 2.11$ with df = 8 and $0.025 < P$-value < 0.05. Logging does significantly reduce the mean number of species in a plot after 8 years at the 5% level, but not at the 1% level. (d) A 90% confidence interval for $\mu_U - \mu_L$ is $(17.5-13.67) \pm 1.860\sqrt{3.53^2/12+4.5^2/9}$ = (0.46, 7.21). (Minitab gives an interval from 0.63964 to 7.02703.) We are 90% confident that the difference in the means for unlogged and logged plots is between 0.46 and 7.21 species.

13.25 Let p_1 denote the proportion of mice ready to breed in good acorn years and p_2 denote the proportion of mice ready to breed in bad acorn years. The sample proportions are $\hat{p}_1 = 54/72 = 0.75$ and $\hat{p}_2 = 10/17 = 0.5882$, and the standard error is $SE = \sqrt{\dfrac{0.75 \times 0.25}{72} + \dfrac{0.5882 \times 0.4118}{17}} \doteq 0.1298$. A 90% confidence interval for $p_1 - p_2$ is $(0.75-0.5882) \pm 1.645 \times 0.1298$ = $(-0.0518, 0.3753)$. With 90% confidence, we estimate that the percent of mice ready to breed in the good acorn years is between 5.2% lower and 37.5% higher than in the bad years. These methods can be used because the populations of mice are certainly more than 10 times as large as the samples, and the counts of successes and failures are at least 5 in both samples. We must view the trapped mice as an SRS of all mice in the two areas.

13.26 (a) The sample proportion of women who felt vulnerable is $\hat{p}_W = \dfrac{27}{56} \doteq 0.4821$, and the corresponding sample proportion for men is $\hat{p}_M = \dfrac{46}{63} \doteq 0.7302$. (b) A 95% confidence interval for the difference $p_M - p_W$ is $(0.7302-0.4821) \pm 1.96\sqrt{\dfrac{0.7302 \times 0.2698}{63} + \dfrac{0.4821 \times 0.5179}{56}} =$ (0.0773, 0.4187). With 95% confidence, we estimate the percent of men who feel vulnerable in this area to be about 0.08 to 0.42 above the proportion of women who feel vulnerable. Notice that 0 is not included in our confidence interval, so there is a significant difference between these proportions at the 5% level.

13.27 (a) A 95% confidence interval for p_N is $\frac{5690}{12931} \pm 1.96\sqrt{\frac{0.44 \times 0.56}{12931}}$ = (0.4315, 0.4486). With 95% confidence, we estimate the percent of cars that go faster than 65 mph when no radar is present is between 43.15% and 44.86%. (b) A 95% confidence interval for $p_N - p_R$ is $(0.44 - 0.32) \pm 1.96\sqrt{\frac{0.44 \times 0.56}{12931} + \frac{0.32 \times 0.68}{3285}}$ = (0.102, 0.138). With 95% confidence, we estimate the percent of cars going over 65 mph is between 10.2% and 13.8% higher when no radar is present compared to when radar is present. (c) In a cluster of cars, where one driver's behavior might affect the others, we do not have independence; one of the important properties of a random sample.

13.28 A 95% confidence interval for p is $\frac{1318}{2092} \pm 1.96\sqrt{\frac{0.63 \times 0.37}{2092}}$ = (0.6093, 0.6507). We are 95% confident that between 61% and 65% of all adults use the internet. (b) A 95% confidence interval for $p_U - p_N$ is $(0.79 - 0.38) \pm 1.96\sqrt{\frac{0.79 \times 0.21}{1318} + \frac{0.38 \times 0.62}{774}}$ = (0.3693, 0.4506). We are 95% confident that the difference in the proportion of internet users and nonusers who expect businesses to have Web sites is between 0.37 and 0.45.

13.29 Let p_1 = the proportion of students who use illegal drugs in schools with a drug testing program and p_2 = the proportion of students who use illegal drugs in schools without a drug testing program. We want to test $H_0: p_1 = p_2$ versus $H_a: p_1 < p_2$. The combined sample proportion is $\hat{p}_c = \frac{7+27}{135+141} \doteq 0.1232$ and the test statistic is

$$z = \frac{0.0519 - 0.1915}{\sqrt{0.1232(1-0.1232)(1/135 + 1/141)}} \doteq -3.53$$

, with a P-value = 0.0002. Since 0.0002 < 0.01, we reject H_0. There is extremely strong evidence that drug use among athletes is lower in schools that test for drugs. There should be some concern expressed about the condition of two independent simple random samples, because these two samples may not be representative of similar schools.

13.30 (a) The patients were randomly assigned to two groups. The first group of 1649 patients received only aspirin and the second group of 1650 patients received aspirin and dipyridamole. (b) We want to test $H_0: p_1 = p_2$ versus $H_a: p_1 \neq p_2$. The combined sample proportion is $\hat{p}_c = \frac{206+157}{1649+1650} \doteq 0.11$ and the test statistic is $z = \frac{0.1249 - 0.0951}{\sqrt{0.11(1-0.11)(1/1649 + 1/1650)}} \doteq 2.73$, with a P-value = 0.0064. Since 0.0064 < 0.01, there is very strong evidence that there is a significant difference in the proportion of strokes between aspirin only and aspirin plus dipyridamole. (c) A 95% confidence interval for $p_1 - p_2$ is

$\left(0.1104-0.1121\right)\pm 1.96\sqrt{\frac{0.1104\times 0.8896}{1649}+\frac{0.1121\times 0.8879}{1650}}$ = (−0.0232, 0.0197). We are 95% confident that the difference in the proportion of deaths for the two treatment groups is between −0.02 and 0.02. Notice that 0 is in the confidence interval, so we do not have evidence of a significant difference in the proportion of deaths for these two treatments at the 5% level. (d) A Type I error is committed if the researchers conclude that there is a significant difference in the proportions of strokes with these two treatments, when in fact there is no difference. A Type II error is committed if the researchers conclude that there is no difference in the proportions of strokes with these two treatments, when in fact there is a difference. A Type II error is more serious because no patients would be harmed with a Type I error, but patients suffer unnecessarily from strokes if the best treatment is not recommended.

13.31 For computer access at home, we want to test $H_0: p_B = p_W$ versus $H_a: p_B \neq p_W$. The combined sample proportion is $\hat{p}_c = \frac{86+1173}{131+1916} \doteq 0.615$ and the test statistic is $z = \frac{0.6565-0.6122}{\sqrt{0.615(1-0.615)(1/131+1/1916)}} \doteq 1.01$, with a P-value = 0.3124. The same hypotheses are used for the proportions with computer access at work. The combined sample proportion is $\hat{p}_c = \frac{100+1132}{131+1916} \doteq 0.602$ and the test statistic is $z = \frac{0.7634-0.5908}{\sqrt{0.602(1-0.602)(1/131+1/1916)}} \doteq 3.90$, with a P-value < 0.0004. Since the P-value is below any reasonable significance level, say 1%, we have very strong evidence of a difference in the proportion of blacks and whites who have computer access at work.

13.32 (a) Let p_1 = the proportion of women got pregnant after in vitro fertilization and intercessory prayer and p_2 = the proportion of women in the control group who got pregnant after in vitro fertilization. We want to test $H_0: p_1 = p_2$ versus $H_a: p_1 \neq p_2$. The combined sample proportion is $\hat{p}_c = \frac{44+21}{88+81} \doteq 0.3846$ and the test statistic is $z = \frac{0.5-0.26}{\sqrt{0.3846(1-0.3846)(1/88+1/81)}} \doteq 3.21$, with a P-value = 0.0014. Since 0.0014 < 0.01, we reject H_0. This is very strong evidence that the observed difference in the proportions of women who got pregnant is not due to chance. (b) This study shows that intercessory prayer may cause an increase in pregnancy. However, it is unclear if the women knew that they were in a treatment group. If they found out that other people were praying for them, then their behaviors may have changed and there could be many other factors to explain the difference in the two proportions. (c) A Type I error would be committed if researchers concluded that the proportions of pregnancies are different, when in fact they are the same. This may lead many couples to seek intercessory prayer. A Type II error would be committed if researchers concluded that the proportions are not different, when in fact they are different. Couples would fail to take advantage of a helpful technique to improve their chances of having a baby. For couples who are interested in having a baby, a Type II error is clearly more serious.

13.33 (a) H_0 should refer to population proportions p_1 and p_2, not sample proportions. (b) Confidence intervals account only for sampling error.

13.34 (a) Let p_1 = the proportion of households where no message was left and contact was eventually made and p_2 = the proportion of household where a message was left and contact was eventually made. We want to test $H_0: p_1 = p_2$ versus $H_a: p_1 < p_2$. The combined sample proportion is $\hat{p}_c = \dfrac{58+200}{100+291} \doteq 0.66$ and the test statistic is $z = \dfrac{0.58-0.687}{\sqrt{0.66(1-0.66)(1/100+1/291)}} \doteq -1.95$, with a P-value = 0.0256. Yes, at the 5% level, there is good evidence that leaving a message increases the proportion of households that are eventually contacted. (b) Let p_1 = the proportion of households where no message was left but the survey was completed and p_2 = the proportion of household where a message was left and the survey was completed. We want to test $H_0: p_1 = p_2$ versus $H_a: p_1 < p_2$. The combined sample proportion is $\hat{p}_c = \dfrac{33+134}{100+291} \doteq 0.427$ and the test statistic is $z = \dfrac{0.33-0.46}{\sqrt{0.427(1-0.427)(1/100+1/291)}} \doteq -2.28$, with a P-value = 0.0113. Yes, at the 5% level, there is good evidence that leaving a message increases the proportion of households who complete the survey. (c) A 95% confidence interval for the difference $p_1 - p_2$ when dealing with eventual contact is (−0.218, 0.003). A 95% confidence interval for the difference $p_1 - p_2$ when dealing with completed surveys is (−0.239, −0.022). Although these effects do not appear to be large, when you are dealing with hundreds (or thousands) of surveys anything you can do to improve nonresponse in the random sample is useful.

13.35 (a) $H_0: p_1 = p_2$ versus $H_a: p_1 > p_2$ where p_1 is the proportion of all HIV patients taking a placebo that develop AIDS and p_2 is the proportion of all HIV patients taking AZT that develop AIDS. The populations are much larger than the samples, and $n_1\hat{p}_c, n_1(1-\hat{p}_c), n_2\hat{p}_c, n_2(1-\hat{p}_c)$ are all at least 5. (b) The sample proportions are $\hat{p}_1 = \dfrac{38}{435} = 0.0874$, $\hat{p}_2 = \dfrac{17}{435} = 0.0391$, and $\hat{p}_c = 0.0632$. The test statistic is $z = \dfrac{0.0874-0.0391}{\sqrt{0.0632(1-0.0632)(1/435+1/435)}} \doteq 2.93$, with a P-value of 0.0017. There is very strong evidence that a significantly smaller proportion of patients taking AZT develop AIDS than if they took a control. (c) Neither the subjects nor the researchers who had contact with them knew which subjects were getting which drug.

13.36 A Type I error would be committed if researchers concluded that the treatment is more effective than a placebo, when in fact it is not. A consequence is that patients would be taking AZT and perhaps suffering from side effects from the medication that is not helpful. A Type II error would be committed if researchers conclude that there is no difference in the success of

AZT and a placebo, when in fact there is a difference. The consequence is that patients would not get the best possible treatment. A Type II error is more serious in this situation because we want patients to get the best possible treatment.

13.37 (a) The number of orders completed in 5 days or less before the changes was $X_1 = 0.16 \times 200 = 32$. With $\hat{p}_1 = 0.16$ and $SE_{\hat{p}} \doteq 0.02592$, the 95% confidence interval for p_1 is (0.1092, 0.2108). (b) After the changes, $X_2 = 0.9 \times 200 = 180$. With $\hat{p}_2 = 0.9$ and $SE_{\hat{p}} \doteq 0.02121$, the 95% confidence interval for p_2 is (0.8584, 0.9416). (c) The standard error of the difference in the proportions is $SE_{\hat{p}_2 - \hat{p}_1} \doteq 0.0335$ and the 95% confidence interval for $p_2 - p_1$ is (0.6743, 0.8057) or about 67.4% to 80.6%. No, the confidence intervals are not directly related. Each interval is based on a different sampling distribution. Properties of the sampling distribution of the difference can be obtained from properties of the individual sampling distributions in parts (a) and (b), but the upper and lower limits of the intervals are not directly related.

13.38 (a) We must have two simple random samples of high-school students from Illinois; one for freshman and one for seniors. (b) The sample proportion of freshman who have used anabolic steroids is $\hat{p}_F = \frac{34}{1679} \doteq 0.0203$. Since the number of successes (34) and the number of failures (1645) are both at least 10, the *z* confidence interval can be used. A 95% confidence interval for p_F is $0.0203 \pm 1.96\sqrt{\frac{0.0203 \times 0.9797}{1679}}$ = (0.0135, 0.0270). We are 95% confident that between 1.35% and 2.7% of high-school freshman in Illinios have used anabolic steroids. (c) The sample proportion of seniors who have used anabolic steroids is $\hat{p}_S = \frac{24}{1366} \doteq 0.0176$. Notice that 0.0176 falls in the 95% confidence interval for plausible values of p_F from part (b), so there is no evidence of a significant difference in the two proportions. The test statistic for a formal hypothesis test is $z = 0.54$ with a *P*-value = 0.59.

13.39 We want to test $H_0: p_1 = p_2$ versus $H_a: p_1 \neq p_2$. From the output, $z = -3.45$ with a *P*-value = 0.0006, showing a significant difference in the proportion of children in the two age groups who sorted the products correctly. A 95% confidence interval for $p_1 - p_2$ is (−0.5025279, −0.15407588). With 95% confidence we estimate that between 15.4% and 50.3% more 6- to 7-year-olds can sort new products into the correct category than 4- to 5-year-olds.

13.40 (a) The two sample proportions are $\hat{p}_W = \frac{6}{53} \doteq 0.1132$ and $\hat{p}_N = \frac{45}{108} \doteq 0.4167$. (b) We want to test $H_0: p_W = p_N$ versus $H_a: p_W \neq p_N$. The combined sample proportion is $\hat{p}_c = \frac{6+45}{53+108} \doteq 0.3168$ and the test statistic is $z = \frac{0.1132 - 0.4167}{\sqrt{0.3168(1-0.3168)(1/53+1/108)}} \doteq -3.89$, with a *P*-value < 0.0002. Since the *P*-value is less than any reasonable significance level, say

1%, we reject H_0. We have very strong evidence that there is a significant difference between the proportions of injured in-line skaters who sustain wrist injuries with and without wrist guards. (These are SRSs of all people injured while in-line skating with and without wrist guards, so we can only make our inference to these populations.) (c) The proportion of nonresponse is 45/206 = 0.2184 or about 21.84%. (d) Yes. Suppose that all 45 people who were not interviewed were injured while wearing wrist guards. (This is unlikely, but we are looking at the extreme case to see if our answer could change.) The proportion of injuries with wrist guards is now $\hat{p}_W = \frac{6+45}{53+45} \doteq 0.5204$. The test statistic would become $z = 1.49$ with a *P*-value of 0.136, which is not significant.

CASE CLOSED!

(1) a) We want to test $H_0: \mu_1 = \mu_2$ versus $H_a: \mu_1 > \mu_2$, where μ_1 is the mean drive-thru service time for McDonald's (Burger King) in 2001 and μ_2 is the mean percent drive-thru service time for McDonald's (Burger King) in a year after the incentive/rewards programs were implemented. (2) Using Minitab with df = 1293, the 95% confidence interval for $\mu_1 - \mu_2$ is $(170.85 - 152.52) \pm 1.9618\sqrt{17.06^2/750 + 16.49^2/596}$ = (16.5274, 20.1326). With 95% confidence we estimate that the average drive-thru service time decreased between 16.5 and 20.1 seconds after the incentive/rewards program was implemented at McDonald's. (3) We want to test $H_0: \mu_M = \mu_T$ versus $H_a: \mu_M > \mu_T$, where μ_M is the mean drive-thru service time for McDonald's in 2004 and μ_T is the mean percent drive-thru service time for Taco Bell in 2004. The test statistic is $t = \frac{152.52 - 148.16}{\sqrt{16.49^2/596 + 18.71^2/590}} \doteq 4.26$ with df = 100 (the largest value below 589 in Table C) or 1162 (with Minitab) and *P*-value < 0.0005. Yes, these data provide extremely strong evidence that drive-thru service times at Taco Bell were significantly faster than those at McDonald's. (4) We want to test $H_0: p_1 = p_2$ versus $H_a: p_1 < p_2$ where p_1 is the proportion of all orders in 2001 that were filled accurately and correct change was given and p_2 is the proportion of orders in 2002 that were filled accurately and correct change was given. The populations are much larger than the samples, and $n_1\hat{p}_c, n_1(1-\hat{p}_c), n_2\hat{p}_c, n_2(1-\hat{p}_c)$ are all at least 5. (b) The sample proportions are $\hat{p}_1 = \frac{730}{890} \doteq 0.8202$, $\hat{p}_2 = \frac{654}{742} \doteq 0.8814$, and $\hat{p}_c = \frac{730+654}{890+742} = 0.848$. The test statistic is $z = \frac{0.8202 - 0.8814}{\sqrt{0.848(1-0.848)(1/890 + 1/742)}} \doteq -3.43$, with a *P*-value of 0.0003. Yes, there was a significant improvement in accuracy between 2001 and 2002. In short, the difference observed from these two independent samples (or something more extreme) would only occur about 3 times in 10,000 trials. We have very convincing evidence that the observed difference is not due to chance, but to some other factor, perhaps better training by the managers! (5) Let p_C denote the proportion of inaccurate for Chick-fil-A in 2002 and p_M denote the proportion inaccurate orders at McDonald's in 2002. A 95% confidence interval for $p_C - p_M$ is $(0.0714 - 0.12) \pm 1.96\sqrt{\frac{0.0714 \times 0.9286}{196} + \frac{0.12 \times 0.88}{750}}$ = (−0.0915, −0.0057). We are 95%

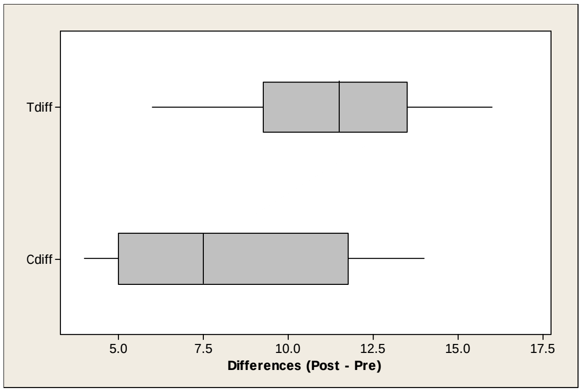
confident that the difference in the proportion of inaccurate orders in 2002 for the two fast food restaurants is between −0.09 and −0.01. Notice that 0 is not in the confidence interval, so there is a significant difference in the proportion of inaccurate orders at the two restaurants.

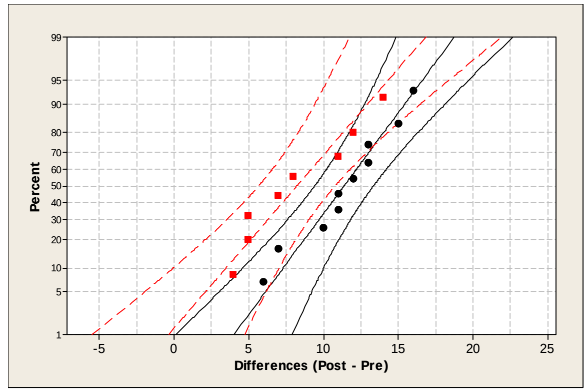
13.41 (a) This is a two-sample *t* test. The two groups of women are (presumably) independent. (b) df = 45 − 1 = 44. (c) The sample sizes are large enough, $n_1 = n_2 = 45$, that the averages will be approximately Normal, so the fact that the individual responses do not follow a Normal distribution has little effect on the reliability of the *t* procedure.

13.42 (a) This is an observational study because the researchers simply observed the random samples of women; they did not impose any treatments. (b) We want to test $H_0: p_N = p_B$ versus $H_a: p_N > p_B$. The combined sample proportion is $\hat{p}_c = \frac{183+68}{220+117} \doteq 0.7448$ and the test statistic is $z = \frac{0.8318-0.5812}{\sqrt{0.7448(1-0.7448)(1/220+1/117)}} \doteq 5.02$, with a *P*-value < 0.0001. We have very strong evidence that a smaller proportion of female Hispanic drivers wear seat belts in Boston than in New York.

13.43 We want to test $H_0: p_H = p_W$ versus $H_a: p_H \neq p_W$. The combined sample proportion is $\hat{p}_c = \frac{286+164}{539+292} \doteq 0.5415$ and the test statistic is $z = \frac{0.5306-0.5616}{\sqrt{0.5415(1-0.5415)(1/539+1/292)}} \doteq -0.86$, with a *P*-value = 0.3898. Since 0.3898 > 0.05, there is not a significant difference between Hispanic and white drivers. For the size of the difference, construct a 95% (or other level) confidence interval. A 95% confidence interval for $p_H - p_W$ is $(0.5306-0.5616) \pm 1.96\sqrt{\frac{0.5306\times 0.4694}{539}+\frac{0.5616\times 0.4384}{292}}$ = (−0.1018, 0.0398). With 95% confidence we estimate the difference in the proportions for Hispanic and white drivers who were seat belts to be between −0.10 and 0.04. Notice that 0 is in the 95% confidence interval, so we would conclude that there is no difference at the 5% significance level.

13.44 We want to test $H_0: \mu_T = \mu_C$ versus $H_a: \mu_T > \mu_C$, where μ_T is the mean difference (post − pre) for the treatment group and μ_C is the mean difference (post − pre) for the control group. The boxplots (on the left below) show that the distributions are roughly symmetric with no outlier, and the Normal probability plots (on the right below) show linear trends which indicate that the Normal distribution is reasonable for these data.





The test statistic is $t = \dfrac{11.40 - 8.25}{\sqrt{3.17^2/10 + 3.69^2/8}} \doteq 1.91$, with $0.025 < P\text{-value} < 0.05$ and df = 7 (Minitab gives a *P*-value of 0.039 with df=13). The *P*-value is less than 0.05, so the data give good evidence that the positive subliminal message brought about greater improvement in math scores than the control. (b) A 90% confidence interval for $\mu_T - \mu_C$ is $(11.40 - 8.25) \pm 1.895\sqrt{3.17^2/10 + 3.69^2/8}$ = (0.03, 6.27) with df = 7; (0.235, 6.065) using Minitab with df = 13. With 90% confidence, we estimate the mean difference in gains to be 0.235 to 6.065 points better for the treatment group. (c) This is actually a repeated measures design, where two measurements (repeated measures) are taken on the same individuals. Many students will probably describe this design as a completely randomized design for two groups, with a twist—instead of measuring one response variable on each individual, two measurements are made and we compare the differences (improvements).

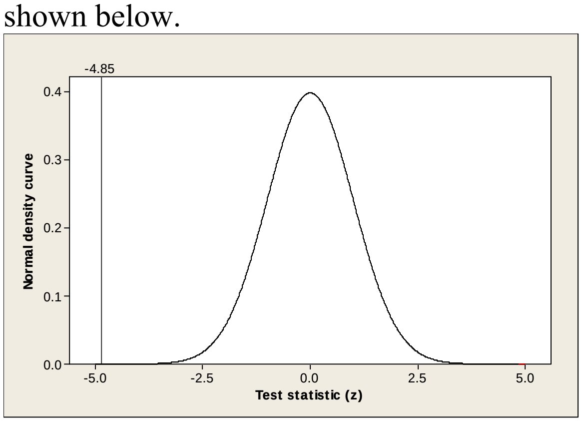
13.45 (a) A 99% confidence interval for $p_M - p_W$ is $(0.9226 - 0.6314) \pm 2.576\sqrt{\dfrac{0.9226 \times 0.0774}{840} + \dfrac{0.6314 \times 0.3686}{1077}}$ = (0.2465, 0.3359). Yes, because the 99% confidence interval does not contain 0. (b) We want to test $H_0: \mu_M = \mu_W$ versus $H_a: \mu_M \neq \mu_W$. The test statistic is $t = \dfrac{272.40 - 274.7}{\sqrt{59.2^2/840 + 57.5^2/1077}} \doteq -0.87$, with a *P*-value close to 0.4. (Minitab reports a P-value of 0.387 with df = 1777.) Since 0.4 > 0.01, the difference between the mean scores of men and women is not significant at the 1% level.

13.46 (a) Matched pairs *t*. (b) Two-sample *t*. (c) Two-sample *t*. (d) Matched pairs *t*. (e) Matched pairs *t*.

13.47 (a) A 99% confidence interval for $\mu_{OPT} - \mu_{WIN}$ is $(7638 - 6595) \pm 2.581\sqrt{289^2/1362 + 247^2/1395}$ = (1016.55, 1069.45). (b) The fact that the sample sizes are both so large (1362 and 1395).

13.48 (a) We want to test $H_0 : \mu_P = \mu_C$ versus $H_a : \mu_P > \mu_C$. The test statistic is $t = \frac{193-174}{\sqrt{68^2/26 + 44^2/23}} \doteq 1.17$, with a P-value close to 0.125. (Minitab reports a P-value of 0.123 with df = 44.) Since 0.125 > 0.05, we do not have strong evidence that pets have higher mean cholesterol than clinic dogs. (b) A 95% confidence interval for $\mu_P - \mu_C$ is $(193-174) \pm 2.074\sqrt{68^2/26 + 44^2/23}$ = (−14.5719, 52.5719). Minitab gives (−13.6443, 51.6443). With 95% confidence, we estimate the difference in the mean cholesterol levels between pets and clinic doges to be between −14 and 53 mg/dl. (c) A 95% confidence interval for μ_p is $193 \pm 2.060\frac{68}{\sqrt{26}}$ = (165.5281, 220.4719). Minitab gives (165.534, 220.466). With 95% confidence, we estimate the mean cholesterol level in pets to be between 165.5 and 220.5 mg/dl. (d) We must have two independent random samples to make the inferences in parts (a) and (b) and a random sample of pets for part (c). It is unlikely that we have random samples from either population.

13.49 (a) The two sample proportions are $\hat{p}_C = 17/283 \doteq 0.0601$ for residents of congested streets and $\hat{p}_B = 35/165 \doteq 0.2121$ for residents of bypass streets. The difference is $\hat{p}_C - \hat{p}_B = -0.1520$ with a standard error of $SE = \sqrt{\frac{0.0601 \times 0.9399}{283} + \frac{0.2121 \times 0.7879}{165}} \doteq 0.0348$. (b) The hypotheses are $H_0 : p_C = p_B$ versus $H_a : p_C < p_B$. The alternative reflects the reasonable expectation that reducing pollution might decrease wheezing. (c) The combined sample proportion is $\hat{p}_c = \frac{17+35}{283+165} \doteq 0.1161$ and the test statistic is $z = \frac{0.0601-0.2121}{\sqrt{0.1161(1-0.1161)(1/283+1/165)}} \doteq -4.85$, with a P-value < 0.0001. A sketch of the distribution of the test statistic, assuming H_0 is true, is shown below.



Notice that a reference line is provided at −4.85 to illustrate how far down in the lower tail of the distribution that this value of the test statistic is located. The P-value tells us the chance of observing a test statistic of −4.85 or something smaller if H_0 is true. As you can see there is almost no chance of this happening, so we have very convincing evidence that the percent of residents reporting improvement from wheezing is higher for residents of bypass streets. (d) The 95% confidence interval, using the standard error from part (b), has margin of error $1.96 \times 0.0348 = 0.0682$. Thus, the 95% confidence interval is -0.152 ± 0.0682 = (−0.2202,

−0.0838). The percentage reporting improvement was between 8% and 22% higher for bypass residents. (e) There may be geographic factors (e.g., weather) or cultural factors (e.g., diet) that limit how much we can generalize the conclusions.

13.50 (a) A 99% confidence interval for $p_H - p_N$ is

$\left(0.07-0.14\right) \pm 2.576\sqrt{\dfrac{0.07\times 0.93}{2455}+\dfrac{0.14\times 0.86}{1191}}$ = (−0.0991, −0.0409). With 99% confidence, the percentage of blacks is between 4.09% and 9.91% higher for non-household providers. Yes, the difference is significant at the 1% level because the 99% confidence interval does not contain 0. (b) A 99% confidence interval for $\mu_H - \mu_N$ is $\left(11.6-12.2\right) \pm 2.581\sqrt{2.2^2/2455+2.1^2/1191}$ = (−0.7944, −0.4056), using df = 1000. (Minitab gives (−0.794182, −0.405818) with df =2456.) With 99% confidence, the mean number of years of school for non-household workers is between 0.41 and 0.79 years higher than household providers. Yes, the difference is significant at the 1% level, because 0 is not included in the 99% confidence interval.