Section 10.1: Categorical Response: Comparing Two Proportions

10.1 Unemployment rate

a) The response variable is unemployment rate and the explanatory variable is race.
b) The two groups that are the categories of the explanatory variable are white and black.
c) The samples of white and black individuals were independent. No individual could be in both samples.

10.2 Sampling sleep

a) The samples on weekdays and weekends should be treated as dependent samples because every person is in both samples.
b) When compared with other people from another year, the samples should be treated as independent. No one person is in both samples.

10.3 Basic life support knowledge

a) The estimated difference between population proportions of students of biological sciences and students of non-biological subjects who know how to activate the MEAS is 0.071. The proportion of students who reported knowing how to activate the MEAS is apparently larger among students of biological sciences than students of non-biological subjects.
b) The standard error is the standard deviation of the sampling distribution of differences between the sample proportions.

$$se = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = \sqrt{\frac{(0.614)(0.386)}{637} + \frac{(0.543)(0.457)}{564}} = 0.0285$$

c) p_1 = proportion of students of biological sciences population who know how to activate the MEAS,

p_2 = proportion of students of non-biological subjects population who know how to activate the MEAS.

d) 0.071 – (1.96)(0.0285) = 0.015

0.071 + (1.96)(0.0285) = 0.127.

(0.015,0.127)

We can be 95% confident that population mean change in proportion is between 0.015 and 0.127. This confidence interval does not contain zero; thus, we can conclude that proportion of students in first-year university students in Brazil who know how to activate the MEAS is apparently larger among students of biological sciences than students of non-biological subjects.

e) The assumptions are that the data are categorical (reported knowing how to activate the MEAS vs. do not know) that the samples are independent and obtained randomly, and there are sufficiently large sample sizes. Specifically, each sample should have at least 10 “successes” and 10 “failures.”

10.4 Less smoking now?

a) The point estimate is 0.541 – 0.238 = 0.303. We estimate that the population percent of adults who never smoked is 30 percentage points higher in the group that did not have any lung obstruction compared to the group that did.
b) We are 99% confident that the population percent of adults who never smoked is between 27 and 34 percentage points higher for adults with no lung obstruction compared to those with lung obstruction.
c) The assumptions are that the response variable is categorical (whether someone never smoked) and observed in two groups (those with and without lung obstruction), that the two samples are independent (the responses from adults without lung obstruction are independent of the ones from adults with lung obstruction) and obtained randomly (given), and that the two sample sizes are sufficiently large (both have at least 10 successes and 10 failures).

10.5 Risky behaviors among HIV positive female sex workers

a) $\hat{p}_1 = 26/76 = 0.3421$, and $\hat{p}_2 = 54/105 = 0.5143$

b) $\hat{p}_1 - \hat{p}_2 = -0.1722$

$$se = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = \sqrt{\frac{0.3421(0.6579)}{76} + \frac{0.5143(0.4857)}{105}} = 0.0721$$

10.5 (continued)

–0.1722 – (1.96)(0.0721) = –0.3143
–0.1722 + (1.96)(0.0721) = –0.0301
(–0.3143, –0.0301)
We can be 95% confident that population proportion for HIV positive female sex workers aged < 35 years who use drug falls between 0.0301 and 0.3143 lower than population proportion for those aged ≥ 35. Because 0 does not fall in this interval, we can conclude that HIV positive females aged ≥ 35 years are more likely than are HIV positive females aged <35 years to use drugs. The assumptions are that the data are categorical, the samples are independent and obtained randomly, and there are sufficiently large sample sizes.

c) The confidence interval has a wide range of plausible values for the population mean difference in proportions, ranging from –0.3143 which represents a fairly large difference to –0.0301 which represents a more modest difference.

10.6 Aspirin and heart attacks in Sweden

a) Each "Sample p" is obtained by taking the proportion of the people in that sample who had a heart attack.
b) The "estimate for p(1) – p(2)" is obtained by subtracting the "Sample p" for the second sample from the "Sample p" from the first sample. There is a difference of 0.014.
c) The confidence interval tells us that we can be 95% confident that the population difference in proportions is between –0.005 and 0.033. Because zero is in this interval, it is plausible that there is no difference between proportions. There may be no difference in proportions of heart attacks between the aspirin and placebo groups.
d) The estimate for the difference would change in sign; it would be negative instead of positive. The endpoints of the confidence interval also would change in signs. They would be (–0.033, 0.005). The confidence interval still includes zero; there may be no difference in proportions of heart attacks between those who take aspirin and those who take placebo.

10.7 Swedish study test

a) $H_0: p_1 = p_2$; $H_a: p_1 \neq p_2$
b) The P-value of 0.14 tells us that, if the null hypothesis were true, we would obtain a difference between sample proportions at least this extreme 0.14 of the time.
c) The bigger the sample size, the smaller the standard error and the bigger the test statistic. This study has smaller samples than the Physicians Health Study did. Therefore, its standard error was larger and its test statistic was smaller. A smaller test statistic has a larger P-value.
d) The P-value would be 0.14/2 = 0.07.

10.8 Significance test for aspirin and heart attacks study

a) $\hat{p} = (347+327)/(11{,}535+14{,}035) = 674/25{,}570 = 0.026$

$$se_0 = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} = \sqrt{0.026(1-0.026)\left(\frac{1}{11{,}535}+\frac{1}{14{,}035}\right)} = 0.002$$

b) $z = \dfrac{(\hat{p}_1-\hat{p}_2)-0}{se_0} = \dfrac{0.0301-0.0233}{0.002} = 3.4$

c) The P-value is 0.001. If the null hypothesis were true, the probability would be close to 0 of getting a test statistic at least as extreme as the value observed. We have strong evidence that there is a difference in the proportion of cancer deaths between those taking placebo and those taking aspirin.

10.9 Basic life support knowledge and enrollment willingness in a first-aid course.

a) Assumptions: Each sample must have at least 10 outcomes of each type. The data must be categorical, and samples must be independent random samples.
Notation: p is the probability that a student says that he/she would enroll in a first aid course to cultivate awareness about basic life support knowledge. p_1 = population proportion of students of biological subjects who would enroll in a first aid course, p_2 = population proportion of students of non-biological subjects who would enroll in a first aid course. $H_0: p_1 = p_2$; $H_a: p_1 \neq p_2$

10.9 (continued)

b) $x_1 = n_1\hat{p}_1 = 637 \times 0.747 = 476$ (students of biological sciences) and $x_2 = n_2\hat{p}_2 = 564 \times 0.511 = 288$ (students of non-biological subjects) $\hat{p} = (476 + 288)/(637 + 564) = 0.636$; this is the common value of p_1 and p_2, estimated by the proportion of the total sample who reported that they would enroll in a first aid course.

c) $se_0 = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} = \sqrt{0.636(0.364)\left(\frac{1}{637}+\frac{1}{564}\right)} = 0.028$. In this case, the standard error is interpreted as the standard deviation of the estimates $(\hat{p}_1 - \hat{p}_2)$ from different randomized studies using these sample sizes.

$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{se_0} = \frac{0.747 - 0.511}{0.028} = 8.43$. *P*-value < 0.0001; if the null hypothesis were true, the probability would be 0.0001 of getting a test statistic at least as extreme as the value observed. We have sufficient evidence to reject the null hypothesis; there is a difference in proportions of reports of willingness to enroll in a first aid course to cultivate awareness about basic life support knowledge among students of biological sciences and non-biological subjects.

10.10 Comparing marketing commercials

a) Here are the results from software:

Sample	X	N	Sample p
1	25	100	0.250000
2	20	100	0.200000

Difference = p (1) – p(2)
Estimate for difference: 0.05
95% CI for difference: (–0.0655382, 0.165538)
Test for difference = 0 (vs. not = 0): Z = 0.85 P-Value = 0.396

1) Assumptions: Each sample must have at least ten outcomes of each type. The data must be categorical, and the samples must be independent random samples.
2) $H_0: p_1 = p_2$; $H_a: p_1 \neq p_2$
3) $z = 0.85$
4) P-value: 0.40
5) If the null hypothesis were true, the probability would be 0.40 of getting a test statistic at least as extreme as the value observed. It is plausible that the null hypothesis is correct, and that there is no population difference in proportions of Group A and B who say they would buy the product.

b) The manager's conclusion is not supported by the data; we do not have strong enough evidence to make this conclusion. One limitation of this study is the volunteer nature of the sample. It is not a random sample.

10.11 Hormone therapy for menopause

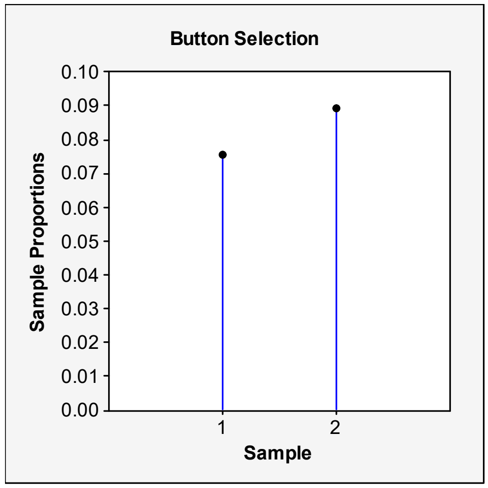
a) Assumptions: Each sample must have at least ten outcomes of each type. The data must be categorical, and the samples must be independent random samples. p is the probability that someone developed cancer. The hypotheses are $H_0: p_1 = p_2$; $H_a: p_1 \neq p_2$.

b) The test statistic is 1.03, and the P-value is 0.303 (rounds to 0.30). If the null hypothesis were true, the probability would be 0.30 of getting a test statistic at least as extreme as the value observed. It is plausible that the null hypothesis is correct and that there are the results for the hormone therapy group are not different from the results for the placebo group.

c) We cannot reject the null hypothesis.

10.12 Obama A/B testing

a)



b) 1) Assumptions: Categorical response (whether visitor clicked button), two groups (two versions of the website), independent and random samples, at least 10 successes and 10 failures in each group.

2) Hypotheses: $H_0: p_1 = p_2$; $H_a: p_1 \neq p_2$, where p_1 and p_2 are the population proportion of visitors who clicked the button on the original and alternative version of the website, respectively.

3) Test Statistic: $z = -10.03$

4) The P-value is approximately 0.

5) Conclusion: With this small P-value, there is strong evidence of a significant difference. The population proportion of visitors who clicked the button differs between those visiting the original site and those visiting the alternative version.

c) The percentage of visitors clicking the button on the original website is at least 1.1 percentage points and at most 1.7 percentage points lower than on the alternative version of the website. This interval tells us about the range of the effect, rather than just telling us that the effect is significant.

10.13 Prevalence of allergen-specific IgE antibodies in schoolchildren

One of the assumptions for the inference in this section is that we have two independent samples. Here, authors surveyed the same subjects twice, so the responses obtained in 2001 are not independent from those obtained in 1996. Therefore, the methods used by the authors are not those of this section (but see Section 10.4).

Section 10.2: Quantitative Response: Comparing Two Means

10.14 Energy drinks: health risks and toxicity

a) The two groups compared were overweight/obese male Saudi university students (Group 1) and normal weights Saudi male university students (Group 2). Let μ_1 be the population mean heart rate range MHRR for overweight/obese students and μ_2 the population mean heart rate range MHRR for normal weights students. $H_0: \mu_1 = \mu_2$

b) A confidence interval estimates a range of possible values for the difference in the population mean heart rate range (MHRR), whereas a *P*-value just indicates whether the difference is statistically significant without telling the size of the effect.

10.15 Address global warming

a) The response variable is the amount of tax the student is willing to add to gasoline in order to encourage drivers to drive less or to drive more fuel-efficient cars; the explanatory variable is whether the student believes that global warming is a serious issue that requires immediate action or not.

b) Independent samples; the students were randomly sampled so which group the student falls in (yes or no to second question) should be independent of the other students.

10.15 (continued)

c) A 95% confidence interval for the difference in the population mean responses on gasoline taxes for the two groups, $\mu_1 - \mu_2$, is given by $(\bar{x}_1 - \bar{x}_2) \pm t_{.025}(se)$ where $\bar{x}_1$ is the sample mean response on the gasoline tax for the group who responded "yes" to the second question, $\bar{x}_2$ is the sample mean response on the gasoline tax for the group who responded "no" to the second question, t is the t-score for a 95% confidence interval and $se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ is the standard error of the difference in mean responses.

10.16 Housework for women and men

a) The study estimated that, on the average, women spend 33.0 – 19.9 = 13.1 more hours than men on housework.

b) The standard error for comparing the means is 1.20. The standard error is so small compared to the sample standard deviations for the two groups because the sample sizes are so large.

$$se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{21.9^2}{476} + \frac{14.6^2}{496}} = 1.20$$

c) The 95% confidence interval is $(\bar{x}_1 - \bar{x}_2) \pm t_{.025}(se) = (33.0 - 19.9) \pm 1.96(1.2)$, or (10.7, 15.5). We can be 95% confident that the difference between the population mean scores of women and men falls between 10.7 and 15.5. Because zero is not in the interval, we can conclude that there is a mean difference between the populations. It appears that the population mean for women is higher than the population mean for men.

d) The assumptions are that the data are quantitative, both samples are independent and random, and there is approximately a normal population distribution for each group.

10.17 More confident about housework

a) The 99% confidence interval is $(\bar{x}_1 - \bar{x}_2) \pm t_{.005}(se) = (33.0 - 19.9) \pm 2.58(1.2)$, or (10.0, 16.2).

b) This interval is wider than the 95% confidence interval because we have chosen a larger confidence level, and thus, the t value associated with it will be higher. To be more confident, we must include a wider range of plausible values.

10.18 Employment by gender

a) It does seem plausible that employment has a normal distribution for each gender because the standard deviations are close in size and much smaller than the means, an indication of a fairly normal distribution.

b) The assumption of an approximately normal population distribution for each group is satisfied and so our inferences are not affected. The remaining assumptions, quantitative response variable for two groups and independent random samples are also satisfied.

c) We can be 95% confident that the population mean gender difference in weekly time spent in employment is between 4.5 and 6.6 hours.

d) The population means are not likely equal. Because 0 is not included in the interval, we can conclude that there is a difference between genders in the population means with respect to time spent in employment. It appears that men spend more time in employment (on the average) than do women.

10.19 Ideal number of children

a) The standard error is 0.043.

$$se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(0.89)^2}{921} + \frac{(0.85)^2}{754}} = 0.043$$

b) The 95% confidence interval is $(\bar{x}_1 - \bar{x}_2) \pm t_{.025}(se) = (2.54 - 2.50) \pm 1.96(0.043)$, or (–0.04, 0.12). In the U.S. population, the mean number of children that women think is ideal is between 0.04 smaller and 0.12 larger compared to what men think is ideal. Because 0 is in the confidence interval, with 95% confidence, the population mean number of children that women and men think is ideal does not differ significantly between the sexes.

10.20 Annual income of CEOs

a) It does not seem plausible that total annual pay of CEOs has a normal distribution for each companies' category because the standard deviations are about the same size as the means indicating skew. Specifically, the lowest possible (but not logical) value of £0 is only 1658/1314 = 1.26 standard deviations below the mean for companies with one compensation consultant and 1779/1461 = 1.22 standard deviations below the mean for companies with two compensation consultants, indicating skewness to the right in both the cases.

b) One of the assumptions for this inference is that both populations are normally distributed. Although we do not likely have normal population distributions, the two-sided test is robust with respect to that assumption, particularly with such large sample sizes. Our inference is not likely affected. The remaining assumptions, quantitative response variable for two groups and independent random samples are satisfied.

c) We can be 95% confident that population mean gender difference in yearly income is between –£370,190 and £128,190. The population means are likely not statistically different. Because £0 falls in that range, we can conclude that there is no difference in the population means with respect to CEOs total annual pay, in UK companies, according to the number of compensation consultants employed by the companies.

10.21 Bulimia CI

a) $se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(2.1)^2}{13} + \frac{(3.2)^2}{17}} = 0.97$

b) The 95% confidence interval is $(\bar{x}_1 - \bar{x}_2) \pm t_{.025}(se) = (2.0 - 4.8) \pm 2.048(0.97)$, or (–4.79, –0.81).
We can be 95% confident that the difference in the population mean family cohesion scores between sexually abused students and non-abused students is between –4.79 and –0.81. Since 0 is not contained in the interval, we can conclude that the mean family cohesion score is lower for the sexually abused students than for the non-abused students.

10.22 Empagliflozin and renal function over time

a) We can be 95% confident that the population mean difference in eGFR between the baseline of the study and four weeks later, was between 0.58 and 0.66 ml per minute per 1.73 m^2 of body-surface area in the 10-mg group

b) The hypotheses are $H_0: \mu_1 = \mu_2$; $H_a: \mu_1 \neq \mu_2$. The *p*-value of <0.001 is too small, and thus provides a statistical evidence to support the alternative hypothesis H_a of a significant difference in the eGFR means between the 10 mg dose Empagliflozin group with the placebo group.

10.23 Nicotine dependence

a) (i) The overwhelming majority of noninhalers must have had HONC scores of 0 because the mean is very close to 0 (and there's a small standard deviation). It would only be this low with a large number of scores of 0.
(ii) On the average, those who reported inhaling had a mean score that was 2.9 – 0.1 = 2.8 (rounds to 3) higher than did those who did not report inhaling.

b) The HONC scores were probably not approximately normal for the noninhalers. The lowest possible value of 0, which was very common, was only a fraction of a standard deviation below the mean.

c) The standard error is interpreted as the standard deviation of the difference between sample means from different studies using these sample sizes.

$$se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(3.6)^2}{237} + \frac{(0.5)^2}{95}} = 0.24$$

d) Because 0 is not in this interval, we can conclude that there is a difference in population mean HONC scores between inhalers and noninhalers at the 95% confidence level. Inhalers appear to have a higher mean HONC score than noninhalers do.

10.24 Inhaling affect HONC?

a) $t = \frac{(2.9 - 0.1) - 0}{0.239} = 11.7$; the P-value associated with this is approximately 0; if the population means were equal, the probability of getting a test statistic this large is about 0.

10.24 (continued)

b) We would reject the null hypothesis. Since the sample mean for inhalers is higher than noninhalers, we can conclude that those who inhaled have a higher population mean HONC score than those who did not inhale.

c) The assumptions are that the data are quantitative, both samples are independent and random, and there is approximately a normal population distribution for each group.

10.25 Females or males more nicotine dependent?

a) The standard error of 0.364 is the standard deviation of the difference between samples from different studies using these sample sizes.

$$se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(3.6)^2}{150} + \frac{(2.9)^2}{182}} = 0.36$$

b) $t = \dfrac{(\bar{x}_1 - \bar{x}_2) - 0}{se} = \dfrac{(2.8 - 1.6) - 0}{0.364} = 3.30$; P-value: 0.001

If the null hypothesis were true, the probability would be 0.001 of getting a test statistic at least as extreme as the value observed. We have very strong evidence that there is a difference between men's and women's population mean HONC scores. The females seem to have a higher population mean HONC score than the males.

c) The HONC scores were probably not normal for either gender. The standard deviations are bigger than the means, an indication of skew. The lowest possible value of 0 is (0 – 2.8)/3.6 = –0.778, or 0.778 standard deviations below the mean for females, and (0 – 1.6)/2.9 = –0.552, or 0.552 standard deviations below the mean for males, indicating skew to the right in both cases. This does not affect the validity of our inference greatly because of the robustness of the two-sided test for the assumption of a normal population distribution for each group.

10.26 Female and male monthly smokers

b) From software: $t = 2.54$; P-value: 0.012. If the null hypothesis were true, the probability would be 0.012 of getting a test statistic at least as extreme as the value observed. We have strong evidence that there is a difference between men's and women's population mean HONC scores among those who were "monthly smokers." The females seem to have a higher population mean HONC score than the males do.

c) The HONC scores were probably not normal for either gender. The standard deviations are almost as large as the means, an indication of skew. The lowest possible value of 0 is (0 – 5.4)/3.5 = –1.543, or 1.543 standard deviations below the mean for females, and (0 – 3.9)/3.6 = –1.083, or 1.083 standard deviations below the mean for males, indicating skew to the right in both cases. Because of the robustness of the two-sided test for the assumption of a normal population distribution for each group, this probably does not have a great effect on the validity of our analysis.

10.27 Kuwaiti men versus Swedish men

a) From software:

Two sample T hypothesis test:
μ_1: Mean of population 1 (Kuwaiti men)
μ_2: Mean of population 2 (Swedish men)
$\mu_1 - \mu_2$: Difference between two means
H_0: $\mu_1 - \mu_2 = 0$
H_A: $\mu_1 - \mu_2 \neq 0$

Difference	Sample Diff.	Std. Err.	DF	T-Stat	P-value
$\mu_1 - \mu_2$	10.84	7.516	22.1	1.4423	0.1646

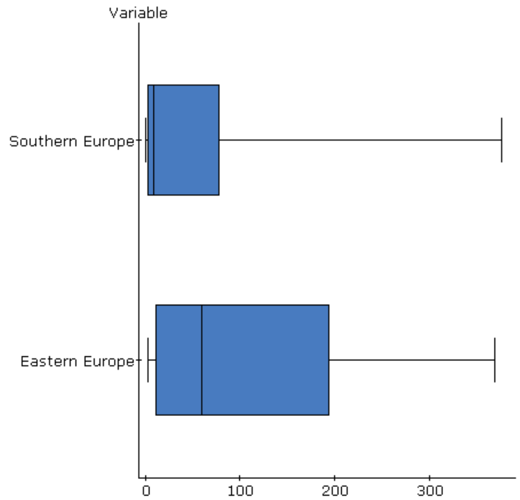
T-Stat = 1.4423; *P*-value = 0.1646. If the null hypotheses were true, the probability would be 0.1646 of getting a test statistic at least as extreme as the value observed. We have strong evidence to not reject the hypothesis of non-difference between Kuwaiti men's and Swedish men's population mean weight. Both populations seem to have a comparable mean weight.

10.27 (continued)

b) The weight distributions are most probably normal; however, the Kuwaiti population seems to be skewed right. The standard deviation for Kuwaiti weights is relatively large compared to the mean, an indication of skew. The lowest possible (but not logical) value of 0 is barely 81.57/26.26 = 3.1 standard deviations below the mean indicating a most likely skew to the right. Because of the robustness of the two-sided test for the assumption of a normal population distribution for each group, this probably does not have a great effect on the validity of this analysis.

10.28 Kidnapping in southern and eastern European countries

a) This plot indicates that the population distributions for both eastern and southern European countries might be skewed to the right.



b) **Two sample T confidence interval:**
μ_1: Mean of eastern European
μ_2: Mean of southern Europe
$\mu_1 - \mu_2$: Difference between two means (with pooled variances)
95% confidence interval results:

Difference	Sample Diff.	Std. Err.	DF	L. Limit	U. Limit
$\mu_1 - \mu_2$	38.92	61.8	14.1	–93.56	171.39

95% CI for difference: (–93.56, 171.39). We can be 95% confident that the population mean difference in kidnapping counts between Eastern and Southern Europe regions is between –93.56 and 171.39. Because 0 is included in this interval, it is plausible that there is no population mean difference in kidnapping counts.

c) **Hypothesis test results:**

Difference	Sample Diff.	Std. Err.	DF	T-Stat	P-value
$\mu_1 - \mu_2$	38.92	61.8	14.1	0.63	0.54

1) The assumptions made by these methods are that the data are quantitative, both samples are independent and random, and there is approximately a normal population distribution for each group.
2) $H_0: \mu_1 = \mu_2$; $H_a: \mu_1 \neq \mu_2$
3) *T*-Stat = 0.63.
4) *P*-value: 0.54.
5) If the null hypothesis were true, the probability would be 0.54 of getting a test statistic at least as extreme as the value observed. It is plausible that the null hypothesis is correct and there is no mean population difference in kidnapping counts between Eastern and Southern Europe regions.

d) The assumptions are listed in part (c). The population distributions may be skewed right. Nevertheless, the two-sided test is robust to violations of the normality assumption.

10.29 Study time

a) Let group 1 represent the students who planned to go to graduate school and group 2 represent those who did not. Then, $\bar{x}_1 = 11.67$, $s_1 = 8.34$, $\bar{x}_2 = 9.10$, and $s_2 = 3.70$. The sample mean study time per week was higher for the students who planned to go to graduate school, but the times were also much more variable for this group.

b) If further random samples of these sizes were obtained from these populations, the differences between the sample means would vary. The standard error of these values would equal about 2.2.

$$se = \sqrt{\frac{8.34^2}{21} + \frac{3.70^2}{10}} = 2.16$$

c) The 95% confidence interval for this data is (–1.9, 7.0). We are 95% confident that the difference in the mean study time per week between the two groups is between –1.9 and 7.0 hours. Since 0 is contained within this interval, we cannot conclude that the population mean study times differ for the two groups.

10.30 Gum flavor longevity

Two sample T hypothesis test:
μ_1: Mean of females
μ_2: Mean of males
$\mu_1 - \mu_2$: Difference between two means
H_0: $\mu_1 - \mu_2 = 0$
H_A: $\mu_1 - \mu_2 \neq 0$ (with pooled variances)
Hypothesis test results:

Difference	Sample Diff.	Std. Err.	DF	T-Stat	P-value
$\mu_1 - \mu_2$	1.34	3.195	12.72	0.42	0.68

1) Assumptions: The data are quantitative (number of minutes of chewing gum flavor longevity); the samples are independent and we will assume that they were collected randomly; we assume that the number of minutes of chewing gum flavor longevity is approximately normal for each group.

2) $H_0: \mu_1 - \mu_2 = 0$; $H_a: \mu_1 - \mu_2 \neq 0$.

3) $t = \dfrac{1.34}{3.195} = 0.42$

4) *P*-value: 0.68

5) If the null hypothesis were true, the probability would be 0.68 of getting a test statistic at least as extreme as the value observed. Since the *P*-value is quite large, we are unable to conclude that there is a significant difference in the average number of minutes of chewing gum flavor longevity for the two groups.

10.31 Time spent on social networks

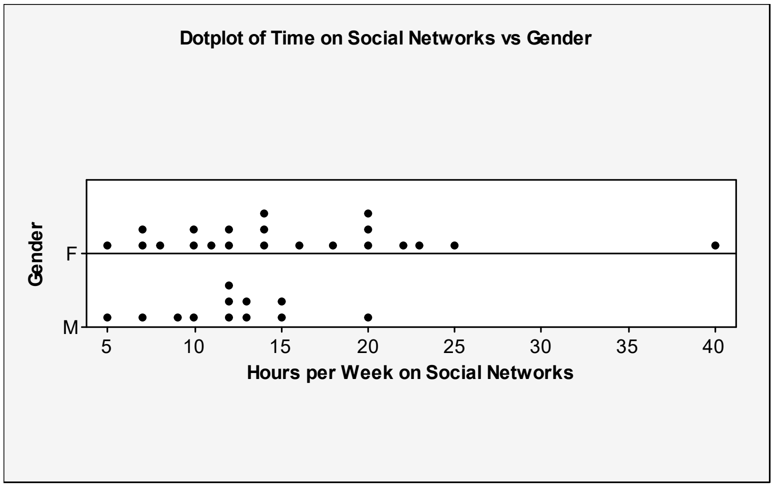
a) Let group 1 represent males and group 2 represent females. Then, $\bar{x}_1 = 11.92$, $s_1 = 3.94$, $\bar{x}_2 = 15.62$, and $s_2 = 8.00$. The sample mean time spent on social networks was higher for females than for males, but notice the apparent outlier for the female group (40). The data were also much more variable for females, but this may also merely reflect the outlier.

b) If further random samples of these sizes were obtained from these populations, the differences between the sample means would vary. The standard deviation of these values would equal about 2.08.

$$se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{3.94^2}{12} + \frac{8.00^2}{21}} = 2.084$$

c) A 90% confidence interval is (–7.24, –0.17). We are 90% confident that the difference in the population mean number of hours spent on social networks per week is between –7.24 and –0.17 for males and females Since 0 is not contained within this interval, we can conclude that the population mean time spent on social networks per week differs for males and females.

10.32 More time on social networks



It appears that the largest value in each group could be an influential outlier. Removing the largest data point in each of the two groups, we obtain $\bar{x}_1 = 11.18$, $s_1 = 3.16$, $\bar{x}_2 = 14.40$, and $s_2 = 5.87$ (much less variability). The sample mean time spent on social networks is still higher for the sample of females. The confidence interval (–5.98, –0.46) still contains values all less than 0.

10.33 Normal assumption

With large random samples, the sampling distribution of the difference between two sample means is approximately normal regardless of the shape of the population distributions. Substituting sample standard deviations for unknown population standard deviations then yields an approximate *t* sampling distribution. With small samples, the sampling distribution is not necessarily bell-shaped if the population distributions are highly non-normal.

10.34 Vital capacity

We can't use the methods from this section because the samples are not independent. Each checkup provided two measurements on the same person, one before and one after using an inhaler. These two measurements are not independent of each other. One can also argue that the biannual measurements are not independent because they are observed on the same person over time.

Section 10.3: Other Ways of Comparing Means, Including a Permutation Test

10.35 Body dissatisfaction

a) The standard error for comparing the means is $s = \sqrt{\dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\dfrac{15(8.0)^2 + 67(6.0)^2}{16+68-2}} =$ 6.413, and the standard error is $se = s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}} = 6.413\sqrt{\dfrac{1}{16} + \dfrac{1}{68}} = 1.782.$

b) The 95% confidence interval is $(\bar{x}_1 - \bar{x}_2) \pm t_{.025}(se) = (13.2 - 7.3) \pm 1.989(1.782)$, or (2.36, 9.44). We can be 95% confident that the population mean difference is between 2.36 and 9.44. Because 0 does not fall in this interval, we can conclude that, on average, the body dissatisfaction assessment score was higher for lean sport athletes than for nonlean sport athletes.

10.36 Body dissatisfaction test

a) $t = \dfrac{(13.2 - 7.3) - 0}{1.782} = 3.311$; P-value: 0.001

b) The assumptions are that the data are quantitative, constitute random samples from two groups, and are from populations with approximately normal distributions. In addition, we assume that the population standard deviations are equal. Given the large standard deviations of the groups, the normality assumption is likely violated, but we're using a two-sided test, so inferences are robust to that assumption.

10.37 Surgery versus placebo for knee pain

a) The confidence interval is (–10.63, 6.43). We can be 95% confident that the population mean pain score is between 10.6 points smaller and 6.4 points larger for patients treated with the placebo procedure compared to patients treated with the lavage procedure. Because 0 falls in this interval, it is plausible that the mean pain score is the same under both procedures.

b) It is reasonable to assume equal population standard deviations, because sample standard deviations s_1 and s_2 are very similar, in fact, identical.

c) 1) We assume independent random samples from the two groups, an approximately normal population distribution for each group (particularly if the sample sizes are small), and equal population standard deviations.

2) $H_0: \mu_1 = \mu_2$; $H_a: \mu_1 \neq \mu_2$

3) $t = -0.49$

4) The P-value is 0.627

5) If there is no difference in the population mean pain score, the probability of observing a test statistic this extreme is 0.63. This is large. There is no evidence of a difference in the population mean pain score between the placebo and lavage arthroscopic surgery procedures.

10.38 Comparing clinical therapies

a) Using technology, choosing "assume equal variances," confirms the given values.

b) We can be 95% confident that the population mean difference between change scores is between –1.2 and 41.2. Because 0 falls in this range, it is plausible that there is no difference in mean change scores between the two populations. We do not have sufficient evidence to conclude that there is a mean difference between the two populations. The therapies may not have different means, but if they do the population mean could be much higher for therapy 1. The confidence interval is so wide because the two sample sizes are very small.

c) Using technology again, the 90% confidence interval is (3.7178, 36.2822). At this confidence level, we can conclude that therapy 1 is better. 0 is no longer in the range of plausible values.

10.39 Clinical therapies 2

a) $H_0: \mu_1 = \mu_2$; $H_a: \mu_1 \neq \mu_2$; $t = ((40.0 - 20.0) - 0)/7.64 = 2.62$; P-Value = 0.059; If the null hypothesis were true, the probability would be 0.059 (rounds to 0.06) of getting a test statistic at least as extreme as the value observed.

b) (i) With a 0.05 significance level, we cannot reject the null hypothesis. It is plausible that the null hypothesis is true and that there is not a population difference in change scores between the two therapy types.

(ii) With a 0.10 significance level, we can reject the null hypothesis. We have strong evidence that there is a difference in population mean change scores between the two therapy types. Therapy 1 led to better results than therapy 2.

c) If the researcher had predicted ahead of time that therapy 1 would be better, that would have corresponded to $H_a: \mu_1 > \mu_2$; the P-value for a one-tailed test would be 0.030. For a significance level of 0.05, we would reject the null hypothesis. We would have strong evidence to conclude that therapy 1 has bigger population change scores than does therapy 2.

10.40 Vegetarians more liberal?

a) The first set of inferences assumes equal population standard deviations, but the sample standard deviations suggest this is not plausible. It is more reliable to conduct the second set of inferences, which do not make this assumption.

b) Based on the first set of results, we would not conclude that the population means are different. 0 falls within the confidence interval; thus, it is plausible that there is no population mean difference. Moreover, the P-value is not particularly small. Based on the second set of results, however, we would conclude that the population means are different. Zero is not in the confidence interval, and hence, is not a plausible value. In addition, the P-value is quite small, an indication that results such as these would be very unlikely if the null hypothesis were true. From this set of results, it appears that the vegetarian students are more liberal than are the non-vegetarian students.

10.41 Teeth whitening results

a) We are testing the hypotheses $H_0: \mu_1 = \mu_2$; $H_a: \mu_1 \neq \mu_2$, where μ_1 represents the population mean change in Vita shade from baseline for the whitening gel group and μ_2 represents the population mean change in Vita shade from baseline for the toothpaste only group. Alternatively, we could test $H_0: \mu_1 - \mu_2 = 0$; $H_a: \mu_1 - \mu_2 \neq 0$.

b) The probability is less than 5% of obtaining a sample difference in means at least as extreme as that observed here assuming the null hypothesis of no difference in population means is true.

c) The pooled standard deviation is $s = \sqrt{\dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\dfrac{57(1.32)^2 + 58(1.29)^2}{58+59-2}} = 1.305$ and the standard error is $se = s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}} = 1.305\sqrt{\dfrac{1}{58} + \dfrac{1}{59}} = 0.241$. The t statistic is $t = \dfrac{\bar{x}_1 - \bar{x}_2}{se} = \dfrac{0.67}{0.241} = 2.78$, $df = n_1 + n_2 - 2 = 58 + 59 - 2 = 115$. The resulting P-value is 0.006. This is statistically significant.

d) The change in mean Vita shade score from baseline was 2.91 times higher for the whitening gel group than for the toothpaste only group.

10.42 Permuting therapies

a)

Patient	All possible assignments					
1	1	1	1	2	2	2
2	1	2	2	1	1	2
3	2	1	2	1	2	1
4	2	2	1	2	1	1

b)

Patient	Score	All possible assignments					
1	30	1	1	1	2	2	2
2	60	1	2	2	1	1	2
3	20	2	1	2	1	2	1
4	30	2	2	1	2	1	1
$\bar{x}_1$		45	25	30	40	45	25
$\bar{x}_2$		25	45	40	30	25	45
$\bar{x}_1 - \bar{x}_2$		20	–20	–10	10	20	–20

10.43 Permutations equally likely

a) The distribution of the improvement scores is the same under therapy 1 and therapy 2.

b) Under H_0, outcomes under each assignment are equally likely. Because there are a total of six possible assignments, each has probability 1/6. Because two assignments lead to a difference of –20, that difference occurs with probability 2/6. Differences –10 and 10 occur once, so each has probability 1/6. The difference of 20 occurs twice, so it has probability 2/6 of occurring.

c) The P-value is the probability of observing a difference as extreme or even more extreme if H_0 is true. The observed difference was 20, the most extreme possible. Because P(20) = 2/6 = 1/3, the P-value is 0.333. If the distribution of improvement scores is the same under therapy 1 and therapy 2, we would observe a difference of 20 or more with a probability of 0.33. This would not be considered unusual, indicating that it is plausible the two therapy distributions of improvement scores (and their means) are the same.

10.44 Two-sided permutation P-value

For the two-sided alternative hypothesis, the extreme differences are the observed difference of 20 in the upper tail and the corresponding difference of –20 in the lower tail of the sampling distribution. These have probabilities of P(20) = 1/3 and P(–20) = 1/3. Therefore, the two-sided P-value is 1/3 + 1/3 = 2/3, or 0.67.

10.45 Time spent on social networks revisited

a) From the app: $\bar{x}_1 = 11.9$, $\bar{x}_2 = 15.6$, and $\bar{x}_1 - \bar{x}_2 = -3.7$.

b Answers will vary. One permutation yielded $\bar{x}_1 = 14.6$, $\bar{x}_2 = 14.1$, and $\bar{x}_1 - \bar{x}_2 = 0.49$.

c) Answers will vary. The one permutation from (b) was less extreme because $-3.7 < 0.49 < +3.7$.

d) Answers will vary slightly. One generation of 10,000 permutations yielded 1498 that resulted in a difference smaller than –3.7 or larger than +3.7 (two-sided alternative).

e) Based on these 10,000 random permutations, the permutation P-value = 1498/10,000 = 0.149. This is not small. If truly the distribution of time spent on social network sites is the same for males and females, then the probability of observing a difference of –3.7 or more extreme (i.e., larger in absolute value) in the sample means is 0.149. This indicates that the null hypothesis is plausible.

10.46 Compare permutation test to *t* test

a) $H_0: \mu_1 = \mu_2$; $H_a: \mu_1 \neq \mu_2$; $t = -1.78$; P-value = 0.085

b) No, the permutation test in Exercise 10.45 with a P-value of 0.149 > 0.10 indicates not enough evidence to reject H_0, whereas the *t*-test with a P-value of 0.085 < 0.10 indicates to reject H_0. The data in both groups indicate a skewed population distribution, which is supported by an outlier in the female group. Also, the sample size in the male group is rather small. The P-value from the permutation approach appears more trustworthy because assumptions for the *t*-test might be violated.

c) Based on 100,000 random permutations, the permutation P-value now equals 0.1098, which is still larger than 0.10. The conclusion to not reject H_0 does not change.

10.47 Dominance of politicians

a) H_0: The population distribution of ratings is the same for clips based on male and female speakers. (This implies the population means are the same.) H_a: The population mean rating is different for clips based on male and female speakers. The observed difference in sample means equals 11.8. Out of 10,000 random permutations, 980 yielded a difference at least as extreme, resulting in a permutation P-value of 0.098. (Your results might differ slightly.) Because 0.098 > 0.05, there is no evidence of a difference in the population mean dominance rating between clips based on female and male speakers. It is plausible that the distributions are the same.

b) From technology: $t = 1.68$, $df = 57.82$, P-value = 0.097 > 0.05. The results are comparable. The sample size is fairly large (30 in each group), and the histogram of the ratings in each group does not indicate major deviations from normality.

10.48 Sampling distribution of $\bar{x}_1 - \bar{x}_2$

a) For large sample sizes, the sampling distribution is approximately normal, with a mean of $\mu_1 - \mu_2$ and a standard error of $\sqrt{s_1^2/n_1 + s_2^2/n_2}$.

b) The sampling distribution derived via the permutation approach has broader tails, showing more variability compared to the normal distribution.

c) The P-value computed from the permutation sampling distribution would be noticeably larger than the one computed from the approximate normal distribution. The area to the right of 100 in the upper tail is noticeably larger under the permutation distribution.

Section 10.4: Analyzing Dependent Samples

10.49 Does exercise help blood pressure?

a) The three "before" observations and the three "after" observations are dependent samples because the same patients are in both samples.

b) The sample mean of the "before" scores is 150, of the "after" scores is 130, and of the difference scores is 20. The difference between the means for the "before" and "after" scores is the same as the mean of the difference scores.

c) From technology, the standard deviation of the difference scores is 5.00. The standard error is $se = s_d/\sqrt{n} = 5/\sqrt{3} = 2.887$; The 95% confidence interval is $\bar{x}_d \pm t_{.025}(se) = 20 \pm 4.303(2.887)$, or (7.6, 32.4). We can be 95% confident that the difference between the population means is between 7.6 and 32.4. Because 0 is not included in this interval and because all differences are positive, we can conclude that there is a decrease in blood pressure after patients go through the exercise program.

10.50 Test for blood pressure

a) $H_0: \mu_1 = \mu_2$; $H_a: \mu_1 \neq \mu_2$

b) The P-value of 0.02 indicates that if the null hypothesis were true, the probability would be 0.02 of getting a test statistic at least as extreme as the value observed. The exercise program does seem beneficial to lowering blood pressure.

c) The assumptions on which this analysis is based are that the sample of difference scores is a random sample from a population of such difference scores and that the difference scores have a population distribution that is approximately normal, particularly if the sample is small ($n < 30$).

10.51 Social activities for students

a) To compare the mean movie attendance and mean sports attendance using statistical inference, we should treat the samples as dependent; the same students are in both samples.

b) There is quite a bit of spread, but the outliers appear in both directions.

c) The 95% confidence interval is $\bar{x}_d \pm t_{.025}(se) = 4 \pm 2.262(5.11)$, or (–7.56, 15.56), $df = 10 - 1 = 9$.

d) The test statistic is $t = \frac{\bar{x}_d - 0}{se} = \frac{4.0 - 0}{5.11} = 0.78$. The P-value is 0.45. It is plausible that the null hypothesis is true. Because the P-value is large, we cannot conclude that there is a population mean difference in attendance at movies versus sports events.

10.52 More social activities

a) We can be 95% confident that the population mean difference score is between –3.3 and 28.9. Because 0 falls in this interval, it is plausible that there is no population mean difference between attendance at parties and sporting events.

b) $H_0: \mu_1 = \mu_2$; $H_a: \mu_1 \neq \mu_2$, the P-value of 0.11 indicates that if the null hypothesis were true, the probability would be 0.11 of getting a test statistic at least as extreme as the value observed. It is plausible that the null hypothesis is true. We cannot conclude that there is a population mean difference in attendance at parties versus sporting events.

c) When we cannot reject the null hypothesis, the confidence interval will include 0 (an indication that it is plausible that there is no population mean difference).

d) The assumptions on which this analysis is based are that the sample of difference scores is a random sample from a population of such difference scores and that the difference scores have a population distribution that is approximately normal, particularly if the sample is small ($n < 30$).

10.53 Movies versus parties

a) 1) The assumptions made by these methods are that the difference scores are a random sample from a population distribution that is approximately normal.

2) $H_0: \mu_1 = \mu_2$; (or the population mean of difference scores is 0); $H_a: \mu_1 \neq \mu_2$

3) $t = -1.62$

4) The P-value is 0.140.

5) If the null hypothesis were true, the probability would be 0.14 of getting a test statistic at least as extreme as the value observed. Because the probability is 0.14 of observing a test statistic or one more extreme by random variation, we have insufficient evidence that there is a mean population difference between attendance at movies versus attendance at parties.

b) The 95% confidence interval is (–21.111, 3.511) which rounds to (–21.1, 3.5); we are 95% confident that the population mean number of times spent attending movies is between 21.1 less and 3.5 higher than the population mean number of times spent attending parties.

10.54 Midsized cars gas mileage

a) A point estimate of the mean mpg change between the overall performance and the in-city performance is 30.4 – 21.47 = 8.93 mpg.

b) This is not sufficient information to find a confidence interval or conduct a test about the change in the mean. We would need to know the difference score for each car model so that we could get the standard deviation of the difference scores and then the standard error of the mean difference.

10.55 Midsized cars gas mileage change

a) The standard deviation of the change in the miles per gallon (mpg) performance could be much smaller because even though there is a lot of variability among the initial and final mpg values of the car models, most car models do not see a large difference in these values over the course of the study, so the mpg changes would not vary a lot.

b) $se = s_d / \sqrt{n} = 1.79/\sqrt{15} = 0.4622$; A 95% confidence interval given by $\bar{x}_d \pm t_{.025}(se)$ has lower endpoint 8.93 – (1.98)(0.4622) = 8.015 and upper endpoint 8.93 + (1.98)(0.4622) = 9.845. Thus, the confidence interval is (8.015, 9.845). 10 is not a plausible mpg change in the population of midsized cars. The plausible mpg change falls in the range from 8.015 to 9.845 mpg.

c) The data must be quantitative, the sample of difference scores must be a random sample from a population of such mpg differences, and the differences in mpg must have a population distribution that is approximately normal (particularly with samples of size less than 30).

10.56 Internet book prices

a) The samples are dependent because they are the prices of the same ten books at two different internet sites.

b) Let group 1 be the prices from Site A and group 2 be the prices from Site B. Then, $\bar{x}_1 =$ \$87.30, $\bar{x}_2 =$ \$83.00, and $\bar{x}_d =$ \$4.30. The sample mean price for the books from Site A is higher than the sample mean price for the books from Site B. Thus, the sample mean of the difference between the prices from these two sites is positive.

c) From technology, the 90% confidence interval for $\mu_1 - \mu_2$ is given by (1.57, 7.03). Since 0 is less than the values in the confidence interval, we can conclude that the prices for textbooks used at her college are more expensive at Site A than at Site B.

10.57 Comparing book prices 2

1) Assumptions: The differences in prices are a random sample from a population that is approximately normal.

2) $H_0: \mu_d = 0$; $H_a: \mu_d \neq 0$

3) $t = \dfrac{\bar{d} - 0}{s_d / \sqrt{n}} = \dfrac{4.3}{4.7152/\sqrt{10}} = 2.88$

4) The P-value is 0.02.

5) If the null hypothesis is true, the probability of obtaining a difference in sample means as extreme as that observed is 0.02. We would reject the null hypothesis and conclude that there is a significant difference in prices of textbooks used at her college between the two sites for $\alpha = 0.05$ or $\alpha = 0.10$, but not for $\alpha = 0.01$.

10.58 Lung capacity revisited

a) $\bar{x}_d = \dfrac{0.28 + (-0.01) + 0.30 + 0.23 + 0.24}{5} = 0.21$ liters

b) From technology, the 95% confidence interval is (0.05, 0.36). With 95% confidence, the population mean FVC of the author improves by at least 0.05 liters and at most 0.36 liters after using the inhaler.

c) From technology, analyzing the data as if we had two independent samples results in a confidence interval of (–0.36, 0.77). Then, we would conclude, with 95% confidence, that the population mean FVC could be as much as 0.36 liters smaller and as much as 0.77 liters larger after using the inhaler. Because 0 is in the confidence interval, it is plausible that there is no improvement, which differs from the conclusion in (b), which showed an improvement of at least 0.05 liters.

10.59 Comparing speech recognition systems

a) $\hat{p}_1 = 1979/2000 = 0.9895$ and $\hat{p}_2 = 1937/2000 = 0.9685$

b) We are 95% confident that the population proportion of correct results for GDMS is between 0.013 and 0.029 higher than the population proportion of correct results for CDHMM.

10.60 Treat juveniles as adults?

a) They are dependent since matched pairs were formed by matching certain criteria and one juvenile from each pair was assigned to either juvenile court or adult court.

b) Let group 1 represent the juveniles assigned to adult court and group 2 represent the juveniles assigned to juvenile court, then $\hat{p}_1 = 673/2097 = 0.32$ and $\hat{p}_2 = 448/2097 = 0.21$.

c) 1) Assumptions: the difference in re-arrest rates are a random sample from a population that is approximately normal.

2) $H_0: \mu_d = 0$; $H_a: \mu_d \neq 0$

3) $z = \frac{b-c}{\sqrt{b+c}} = \frac{515-290}{\sqrt{515+290}} = 7.9$

4) The P-value is approximately 0.

5) If the null hypothesis is true, the probability of obtaining a difference in sample means as extreme as that observed is very close to 0. There is extremely strong evidence of a population difference in the re-arrest rates between juveniles assigned to adult court and those assigned to juvenile court.

10.61 Change coffee brand?

a) The point estimates for the population proportions choosing Sanka for the first and second purchases, respectively, are (i) 204/541 = 0.38 and (ii) 231/541 = 0.43. The estimated difference of population proportions is the difference between the sample means at the two times: 0.43 – 0.38 = 0.05. We estimate the population proportion of people buying Sanka coffee after the advertising campaign was 0.05 larger than before the campaign. In terms of percentages, we estimate 5% more people in the population bought Sanka coffee after the campaign.

b) The confidence interval of (0.01, 0.09) tells us that we can be 95% confident that the proportion of people buying Sanka coffee after the advertising campaign is between 0.01 and 0.09 larger than before it. This means that we are 95% confident that between 1% and 9% more people in the population bought Sanka coffee after the advertising campaign.

c) $H_0: p_1 = p_2$, where p_1 and p_2 are the (marginal) population proportions of people buying Sanka coffee before and after the advertising campaign, respectively. We have sufficient evidence (P-value of 0.02 < 0.05 significance level) that the population proportion of people buying Sanka coffee differs before and after the advertising campaign.

10.62 President's popularity

a)

	This Month	
Last Month	Yes	No
Yes	450	60
No	40	450

b) (i) The sample proportion giving a favorable rating last month is $(450+60)/1000 = 0.51$.

(ii) The sample proportion giving a favorable rating this month is $(450+40)/1000 = 0.49$.

c) The difference is 0.51 – 0.49 = 0.02. For the 1000 subjects surveyed, the proportion giving a favorable rating dropped by 2 percentage points from last month to this month.

d) $z = \frac{b-c}{\sqrt{b+c}} = \frac{40-60}{\sqrt{40+60}} = -2.0$; The P-value is 0.046.

If the null hypothesis were true, the probability would be 0.046 of getting a test statistic at least as extreme as the value observed. We have reasonably strong evidence that there is a difference in the population proportions of people who thought the president is doing a good job between last month and this month.

10.63 Marital Status and Life insurance

a) The point estimate for the difference between the population proportions who are married and who have a life insurance policy is 0.42 – 0.326 = 0.094.

b) (i) The assumptions are that the data are categorical, the samples are dependent and random, and the sum of the two off-diagonal counts in the test is at least 30.

10.63 (continued)

(ii) H_0: $p_1 = p_2$; H_a: $p_1 \neq p_2$

(iii) $z = \dfrac{(\hat{p}_1 - \hat{p}_2) - 0}{se_0} = \dfrac{b-c}{\sqrt{b+c}} = \dfrac{75-47}{\sqrt{75+47}} = 2.535$

(iv) The P-value is 0.011.

(v) This is a small P-value; if the null hypothesis were true, the probability would be 0.011 of getting a test statistic at least as extreme as the value observed. We have strong evidence that there is a difference between the population proportion of married adult males and the population proportion of adult males who have a life insurance policy. It appears that more people are married than having a life insurance policy.

10.64 Marital Status and Life insurance by age

Given the sample's size for a certain range of age and the proportions of respondents who answered yes for each question do not allow us to compare the two proportions inferentially. Because we do not know the off diagonal measures that are the number of respondents who answered "yes" for one question and "no" for the other; needed to compare proportions with dependent samples.

Section 10.5: Adjusting for the Effects of Other Variables

10.65 Benefits of drinking

a) This refers to an analysis of three variables, a response variable, an explanatory variable, and a control variable. The response variable is "whether or not at risk for cardiovascular disease;" the explanatory variable is "whether drink alcohol moderately;" and the control variables are socioeconomic status and mental and physical health.

b) There is a stronger association between drinking alcohol and its effect on risk for cardiovascular disease for subjects who have a higher socioeconomic status.

10.66 Death penalty in Kentucky

a) White victim:

31/391 = 0.079 or 7.9% of white defendants received the death penalty.

7/57 = 0.123 or 12.3% of black defendants received the death penalty.

Black victim:

0/18 = 0.000 or 0.0% of white defendants received the death penalty.

2/108 = 0.019 or 1.9% of black defendants received the death penalty.

These results suggest that, for each type of victim's race, black defendants were more likely to receive the death penalty than white defendants were.

b) The response variable is verdict, the explanatory variable is defendant's race, and the control variable is victim's race.

c)

	Death Penalty		
Defendant's Race	**Yes**	**No**	**Total**
White	31	378	409
Black	9	156	165

The overall proportions of whites and blacks who receive the death penalty, regardless of victim's race, are 0.076 and 0.055, respectively.

d) The data do satisfy Simpson's paradox in that the direction of association reverses when an additional variable is controlled for. This occurs because of the preponderance of whites who are accused of killing whites, a victim group that leads to higher death penalty rates.

e) (i) A test for two proportions could be conducted using the values in (c).

(ii) A test for two proportions could be conducted using proportions calculated from the accompanying table.

10.67 Stress at Work

a) Based on the following tables (the first one is without control variable and the second one is with control variable), we can infer that the overall proportions of females and males who declared having too much stress at work, regardless of the type of working area, are 0.653 and 0.645, respectively. However, if we take the control variable into consideration, these proportions of males and females are, respectively, 0.813 and 0.725 in the intermediate area, 0.455 and 0.65 in the rural area; and 0.636 and 0.568 in the urban area. The results do not illustrate Simpson's paradox in that the direction of no association reverses in stress feeling at work when an additional variable (the working area) is added for control.

		Stress at Work					
		No		Yes		Total	
		Count	Row N %	Count	Row N %	Count	Row N %
Gender	Female	17	34.7%	32	65.3%	49	100.0%
	Male	44	35.5%	80	64.5%	124	100.0%

				Stress at Work					
				No		Yes		Total	
				Count	Row N %	Count	Row N %	Count	Row N %
Type of working area	Intermediate	Gender	Female	3	18.8%	13	81.3%	16	100.0%
			Male	11	27.5%	29	72.5%	40	100.0%
	Rural	Gender	Female	6	54.5%	5	45.5%	11	100.0%
			Male	14	35.0%	26	65.0%	40	100.0%
	Urban	Gender	Female	8	36.4%	14	63.6%	22	100.0%
			Male	19	43.2%	25	56.8%	44	100.0%

b) It appears from the controlled results that rural area is the most appropriate for female workers because this is the only area where the percentage of stressed women is less than the percentage of stress men at work. In addition, the intermediate zone looks like to be the worst working zone for both genders as the percentages of stressed workers will be the highest for both the genders.

10.68 Teacher salary, gender, and academic rank

a) (i) Overall, the difference is 84.4 – 68.8 = 15.6. The mean salary for male faculty was \$15,600 higher than for female faculty.

(ii) The difference was 109.2 – 96.2 = 13 for professors, 77.8 – 72.7 = 5.1 for associate professors, 66.1 – 61.8 = 4.3 for assistant professors, and 46.0 – 46.9 = –0.9 for instructors.

b) Perhaps the proportion of the faculty who are men is relatively higher at the higher academic ranks, for which the salaries are higher, and relatively lower at the lower academic ranks, for which the salaries are lower.

10.69 Family size in Canada

a) The mean number of children for English-speaking families (1.95) is higher than the mean number of children in French-speaking families (1.85).

b) Controlling for province, this association reverses. In each case, there's a higher mean for French-speaking families. For Quebec, the mean number for French-speaking families (1.80) is higher than for English-speaking families (1.64). Similarly, for other provinces, the mean for French-speaking families (2.14) is higher than the mean for English-speaking families (1.97).

c) This paradox likely results from the fact that there are relatively more English-speaking families in the "other" provinces that tend to produce more children regardless of language, and more French-speaking families in Quebec where they tend to have fewer children regardless of language. This illustrates Simpson's paradox.

10.70 Heart disease and age

This occurs because there are more young people, who are at the lower levels of heart disease fatalities, in Utah, and more old people, who are at the higher levels of heart disease, in Colorado. Even though the rates are lower in Colorado at each age level, the heart disease death rates for young people in Utah are still lower than the death rates for old people in Colorado.

10.71 Breast cancer over time

There could be no difference in the prevalence of breast cancer now and in 1900 for women of a given age. Overall, the breast cancer rate would be higher now, because more women live to an old age now, and older people are more likely to have breast cancer.

Chapter Problems: Practicing the Basics

10.72 Pick the method

a) The statistical method would be a significance test, the samples are independent, the relevant parameter is the mean difference between the means, and the inference method would include the calculation of a t statistic and a P-value. The P-value would be compared to a significance level.

b) The statistical method could be either a significance test (significantly different from a mean change of 0?) or a confidence interval (is 0 included in the interval?), the samples are dependent, the relevant parameter is the mean difference score, and the inference method would include either the calculation of a t statistic and P-value, or the calculation of a confidence interval.

c) The statistical method could be either a McNemar test, a significance test comparing proportions from dependent samples, or a confidence interval (is 0 included in the interval?), the samples are dependent, the relevant parameter is the difference between proportions, and the inference method would include either the calculation of a z statistic and P-value using a McNemar test, or the calculation of a confidence interval.

d) The statistical method could be either a significance test (significantly different from a difference in means of 0?) or a confidence interval (is 0 included in the interval?), the samples are independent, the relevant parameter is the difference between means, and the inference method would include either the calculation of a t statistic and P-value, or the calculation of a confidence interval.

10.73 Public versus scientists' opinions on fracking

a) The response variable the opinion on fracking (favor or oppose), and the explanatory variable is the type of survey (U.S. adults/scientists).

b) The separate samples of subjects should be treated as independent sample in order to conduct inference because each survey uses different subjects. (Highly unlikely that a subject from the scientists' survey was also included in the Pew Research Center survey.)

c) Now, the two samples include the same subjects, and so should be treated as dependent samples (matched pairs) because the same scientists who were asked about fracking were asked about offshore drilling.

10.74 BMI then and now

a) The point estimate for the change in the population proportion is $0.69 - 0.66 = 0.03$. It estimates that the population proportion of adults that are overweight increased by 0.03. In terms of percentages, the estimated population percent of adults that are overweight increased by 3 percentage points between 2003/2004 and 2011/2012.

b) The standard error is so small mainly because of the large sample sizes for both samples.

c) With 95% confidence, the population proportion of adults overweight in 2011/2012 is by at least 0.01 and at most 0.05 larger than in 2003/2004. In terms of percentages, with 95% confidence, the population percent of adults overweight in 2011/2012 is by at least 1 and at most 5 percentage points larger than in 2003/2004. Zero is not contained in the interval as a plausible value, indicating the population proportion of adults overweight is greater in 2011/2012 than in 2003/2004.

10.75 Marijuana and gender

a) We are 95% confident that the population proportion of females who have used marijuana is at least 0.0077 lower and at most 0.0887 lower than the population proportion of males who have used marijuana. Because 0 is not in the confidence interval, we can conclude that females and males differ with respect to marijuana use.

b) The confidence interval would change only in sign. It would now be (0.0077, 0.0887). We are 95% confident that the population proportion of males who have used marijuana is at least 0.0077 higher and at most 0.0887 higher than the population proportion of females who have used marijuana.

10.76 Gender and belief in afterlife

a) The sample proportions who report that they believe in an afterlife for females is $1026/1233 = 0.8321$, for males is $757/1009 = 0.7502$, and for the difference between females and males is $0.8321 - 0.7502 = 0.0819$.

10.76 (continued)

b) The standard error for the estimate of $p_1 - p_2$ is $se = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} =$

$\sqrt{\frac{0.8321(1-0.8321)}{1233} + \frac{0.7502(1-0.7502)}{1009}} = 0.017.$ This expresses how much, for samples of these sizes, the difference in the sample proportion varies (roughly on average) around the true (unknown) difference in the population.

c) The 95% confidence interval is $(\hat{p}_1 - \hat{p}_2) \pm z_{.05}(se) = (0.8321 - 0.7502) \pm 1.96(0.017)$, or (0.05, 0.12). Because 0 is not in the confidence interval, we can conclude that the population proportion believing in an afterlife is larger for females.

d) The difference between these population proportions, 0.81 – 0.72 = 0.09, is in the confidence interval. The confidence interval in (c) contains the parameter it is designed to estimate.

10.77 Belief depend on gender?

a) $\hat{p} = \frac{1026+757}{1233+1009} = 0.795$ and

$$se_0 = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{0.795(1-0.795)\left(\frac{1}{1233} + \frac{1}{1009}\right)} = 0.01714$$

b) $z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{se_0} = \frac{0.0819}{0.01714} = 4.78;$ The P-value is approximately 0.

If the null hypothesis were true, the probability would be approximately 0 of getting a test statistic at least as extreme as the value observed. Therefore, we reject the null hypothesis and conclude that the population proportions believing in an afterlife are different for females and males.

c) If the population difference were 0.81 – 0.72 = 0.09, our decision would have been correct.

d) The assumptions on which the methods in this exercise are based are that we used independent random samples for the two groups (okay since from GSS) and that we had at least 5 successes and 5 failures in each sample.

10.78 Females or males have more close friends?

a) The point estimate of the difference between the population means for males and females is 8.9 – 8.3 = 0.6.

b) We are 95% confident that the population mean for males is between 0.60 – 2.1 = –1.5, or 1.5 lower and 0.60 + 2.1 = 2.7, or 2.7 higher than the population mean for females. Because 0 falls in this interval, it is plausible that there is no difference between the population means for males and females.

c) For each gender, it does not seem like the distribution of number of friends is normal. The standard deviations are larger than the means; the lowest possible value of 0 is (0 – 8.3)/15.6 = –0.532, or 0.532 standard deviations below the mean for females and (0 – 8.9)/15.5 = –0.574, or 0.574 standard deviations below the mean for males, an indication of right skew. The confidence interval in (b) is based on the assumption of normal population distributions. The *t*-distribution, however, is robust with two-sided confidence intervals. With large samples ($n > 30$), it should not affect the validity.

10.79 Heavier horseshoe crabs more likely to mate?

a)

The female crabs have a higher median and a bigger spread if they had a mate than if they did not have a mate. The distribution for the female crabs with a mate is right-skewed, whereas the distribution for the female crabs without a mate is symmetrical.

b) The estimated difference between the mean weights of female crabs who have mates and who do not have mates is 2.6 – 2.1 = 0.5.

c) $se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{0.36}{111} + \frac{0.16}{62}} = 0.076$

d) The 90% confidence interval is $(\bar{x}_1 - \bar{x}_2) \pm t_{.05}(se) = 0.5 \pm 1.645(0.076)$, or (0.375, 0.625). Because n_1 and n_2 are large, we approximate $t_{.05}$ with the normal distribution using $z_{.05} = 1.645$. We can be 90% confident that the difference between the population mean weights of female crabs with and without a mate is between 0.375 and 0.625 kg. Because 0 does not fall in this interval, we can conclude that female crabs with a mate weigh more than do female crabs without a mate.

10.80 TV watching and race

a) TV distribution does not likely have a normal distribution for either race because the standard deviations are almost as large as the means. In fact, the lowest possible value of 0 is (0 – 3.97)/3.54 = –1.12, or 1.12 standard deviations below the mean for blacks, and (0 – 2.77)/2.25 = –1.23, or 1.23 standard deviations below the mean for whites indicating skew to the right in both cases. Although inferences comparing population means assume a normal population distribution, the large sample sizes indicate that these skewed distributions do not likely affect the validity of inferences drawn from these samples.

b) We can be 95% confident that the difference between the population means of blacks and whites is between 0.75 hours and 1.65 hours. Because 0 does not fall in this interval, we can conclude that blacks watch more television than whites do, on the average.

c) This inference is based on the assumptions that the data are quantitative, the samples are independent and random (okay since using GSS), and the population distributions for each group are approximately normal, (which is not so important because of large sample size).

10.81 Test TV watching by race

a) $H_0: \mu_1 = \mu_2$; $H_a: \mu_1 \neq \mu_2$

b) The test statistic is 5.25, and the P-value is 0.000. If the null hypothesis were true, the probability would be almost 0 of getting a test statistic at least as extreme as the value observed.

c) Using a significance level of 0.05, we can reject the null hypothesis. We have strong evidence that there is a difference in mean TV watching time between the populations of blacks and whites. Blacks seem to watch more TV than do whites.

d) When we reject the null hypothesis at a significance level of 0.05, a 95% confidence interval does not include the value at the null hypothesis, in this case, 0.

10.82 Ibuprofen and lifespan

a) The response variable is the level of amino acid (quantitative). The explanatory variable is the whether or not the cells received ibuprofen (categorical).

b) The standard error of the mean for the ibuprofen group is $se_1 = s_1/\sqrt{n_1} = 0.0038/\sqrt{6} = 0.0016,$ the standard error of the mean for the untreated group is $se_2 = s_2/\sqrt{n_2} = 0.0050/\sqrt{6} = 0.0020,$ and the standard error for the difference in means is $se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{0.0038^2}{6} + \frac{0.0050^2}{6}} = 0.0026.$

c) With 95% confidence, the population mean level of tryptophan for cells treated with ibuprofen is between 0.007 lower and 0.004 larger compared to untreated cells. Because 0 falls in the interval, it is plausible that the population mean level of tryptophan is the same for cells treated with ibuprofen and untreated cells.

10.83 Time spent on Internet

a) The response variable is the number of hours a week spent on the Internet and is quantitative. The explanatory variable is the respondent's gender and is categorical.

b) The 99% confidence interval is (–0.8, 2.5). With 95% confidence, the population mean number of hours spent on the web is between 0.8 hours shorter and 2.5 hours longer for males compared to females. Because 0 is in the confidence interval, there is no evidence that the population mean time on the web differs by gender.

c)
1) Assumptions: Independent random samples, and number of hours spent on the Internet per week has an approximately normal population distribution for each gender.
2) H_0: $\mu_1 = \mu_2$; H_a: $\mu_1 \neq \mu_2$, where μ_1 is the mean number of hours for the population of all U.S. males and μ_2 is the mean number of hours for the population of all U.S. females.
3) $t = 1.03$.
4) The P-value is 0.3044.
5) If the null hypothesis is true, the probability of obtaining a difference in sample means as extreme as that observed is 0.3044, which is not unusual. At a significance level of 0.05, we do not have evidence to reject H_0. It is plausible that the population mean number of hours a week spent on the Internet is the same for males and females.

10.84 Test–CI connection

Since the 95% confidence interval contains 0, it is plausible that the population mean number of hours a week spent on the Internet is the same for males and females. This is the same conclusion we reached using a hypothesis test with a significance level of 0.05.

10.85 Sex roles

1) Assumptions: The data are quantitative (child's score); the samples are independent and we will assume that they were collected randomly; we assume that the population distributions of scores are approximately normal for each group.
2) $H_0: \mu_1 = \mu_2$; $H_a: \mu_1 \neq \mu_2$, where group 1 represents the group with the male tester and group 2 represents the group with the female tester.
3) $se = \sqrt{\frac{1.4^2}{50} + \frac{1.2^2}{90}} = 0.2349$; $t = \frac{2.9 - 3.2}{0.2349} = -1.2$
4) The P-value is 0.205.
5) If the null hypothesis were true, the probability would be 0.205 of getting a test statistic at least as extreme as the value observed. Since the P-value is quite large, there is not much evidence of a difference in the population mean of the children's scores when the tester is male versus female.

10.86 How often do you feel sad?

a) The P-value of 0.000 indicates that if the null hypothesis were true, the probability would be almost 0 of getting a test statistic at least as extreme as the value observed. In other words, it is extremely unlikely that the population means are equal because, if they were, it is improbable that we'd get a test statistic of 3.88. It appears that women report feeling sad more often than do men, on the average.

b) We can be 95% confident that the difference between the population means for women and men falls between 0.193 and 0.587. Because 0 does not fall in this confidence interval, we can conclude that women report feeling sad more often than do men, on the average. We learn from the confidence interval, but not from the test, the actual range of plausible values for the mean population difference, and we see the difference may be quite small.

c) The assumptions for these analyses are that the data are quantitative, the samples are independent and random, and the population distributions for each group are approximately normal (particularly with small sample sizes). It appears that the population distributions are not normal; the standard deviations are larger than the means. In fact, the lowest possible value of 0 is $(0 – 1.81)/1.98 = –0.914$, or 0.914 standard deviations below the mean for females, and $(0 – 1.42)/1.83 = –0.776$, or 0.776 standard deviations below the mean for males, indicating skew to the right in both cases, an indication of skew. Given the large sample sizes, this does not likely affect the validity of our inferences.

10.87 Parental support and household type

a) The 95% confidence interval is 4 ± 3.4 , or (0.6, 7.4).

b) The conclusion refers to the results of a significance test in that it tells us the P-value of 0.02. If the null hypothesis were true, the probability would be 0.02 of getting a test statistic at least as extreme as the value observed.

10.88 Car bumper damage

Using technology, a 95% confidence interval is (6.2, 14.5). The test statistic is $t = 7.11$ and P-value is 0.003. We can conclude that the population mean is higher for one bumper type. Zero does not fall in the confidence interval, an indication that it is unlikely that there is no population mean difference, and the P-value of 0.003 is quite small, an indication that it would be improbable to obtain a test statistic this large if the null hypothesis is true (and if we're using samples of this size).

10.89 Teenage anorexia

a) The P-value of 0.102 indicates that if the null hypothesis that there is no difference between population mean change scores were true, the probability would be 0.10 of getting a test statistic at least as extreme as the value observed.

b) The assumptions for this analysis are that the data are quantitative, the samples are independent and random, and the population distributions for each group are approximately normal. Based on the box plots (which show outliers and a skew to the right for the cognitive-behavioral group), it would not be a good idea to conduct a one-sided test. It is not as robust as the two-sided test to violations of the normal assumption.

c) The lowest plausible difference between population means is the lowest endpoint of the confidence interval, –0.7, a difference of less than 1 pound. Thus, if there is a change in this direction (cognitive-behavioral group less), then it is less than 1 pound. On the other hand, the highest plausible difference between population means is the highest endpoint of the confidence interval, 7.6. Thus, if there is a change in this direction (cognitive-behavioral group more), then it could be almost as much as 8 pounds.

d) The confidence interval and test give us the same information. We do not reject the null hypothesis that the difference between the population means is 0, and 0 falls in the 95% confidence interval for the difference between the population means.

10.90 Equal pay in sports?

a) The means are 50,637 for males and 38,286 for females with a difference of 50,637 – 38,286 = 12,351.

b) Judging by the dot plots, the population distribution of the prize money earned may be rather right skewed for both male and female skiers (especially males). Further, we are interested in a one-sided test, and the sample size is not that large. All these are conditions under which the *t*-test may not work well.

c) Answers will vary. One generated permutation yielded a difference in the sample means of –12,971.

d) Answers will vary. For our random permutation, –12,971 is less extreme (one-sided test) than the observed difference of 12351.

10.90 (continued)

e) Answers will vary. Of ten generated permutations, five yielded a test statistic at least as extreme. (Make sure you select "greater" for the alternative hypothesis.)
f) Answers will vary. One generation yielded 3249 out of 10,000 permutations resulted in a test statistic at least as extreme.
g) The permutation P-value is 3249/10,000 = 0.325. If the population distribution of the prize money is the same for male and female skiers, observing a difference of 12,351 in our sample occurs with a probability of 0.325. Because this P-value is not small, there is not enough evidence to conclude that the population means of the population distributions differ.
h) No, the shape of the histogram is almost identical, as is the permutation P-value.

10.91 Surgery versus placebo for knee pain

From technology, the 95% confidence interval is (–10.84257, 5.24257). We can be 95% confident that the population mean difference in pain scores falls between –10.8 and 5.2. Because 0 is included in this range, it is plausible that there is no difference between the population mean pain scores of these two groups.

10.92 More knee pain

From technology using $H_0: \mu_d = 0$; $H_a: \mu_d \neq 0$, the t-value is –0.69 and the P-Value is 0.492 with df = 117. The P-value of 0.49 indicates that if the null hypothesis were true, the probability would be 0.49 of getting a test statistic at least as extreme as the value observed. It is plausible that the null hypothesis is true and that there is no difference in population mean pain levels between these two groups. For this inference, we assume quantitative response variables, independent random samples, and approximately normal population distributions for each group.

10.93 Anorexia again

a) We can be 95% confident that the population mean difference is between –0.7 and 7.6. Because 0 falls in this range, it is plausible that there is no population mean difference between these groups.
b) The P-value of 0.10 indicates that if the null hypothesis were true, the probability would be 0.10 of getting a test statistic at least as extreme as the value observed.
c) The P-value is 0.10/2 = 0.05. If the null hypothesis were true, the probability would be 0.05 of getting a test statistic of 1.68 or larger. This P-value is smaller than the one in (b), thus providing stronger evidence against the null hypothesis.
d) For these inferences, we assume a quantitative response variable, independent random samples, and approximately normal population distributions for each group.

10.94 Breast-feeding helps IQ?

a) From technology, the 95% confidence interval is (–13.5414, –6.6586). The confidence interval does not include 0; therefore, we can conclude that there is a difference between the population means. In addition, the P-value (for $t = -5.77$) is almost 0; if the null hypothesis were true, the probability would be almost 0 of getting a test statistic at least as extreme as the value observed. We have strong evidence that mean IQ is higher among the population of babies who have been breast-fed for 7 to 9 months than among the population of babies who had been breast-fed for no longer than a month.
b) This was an observational study because babies were not randomly assigned to condition. There are several potential lurking variables. Parents' education levels or IQ might affect babies' IQ's and might be related to tendency to breast-feed.

10.95 Australian cell phone use

a) The observations are paired because two observations were made on the same driver: were they using their cell phone when the crash occurred and had they used their cell phone at an earlier time when no accident occurred. Thus, methods for dependent samples should be used.
b) McNemar's test is used for comparing (marginal) proportions from matched pairs.

10.96 Improving employee evaluations

We would explain that there's less than a 5% chance that we'd get a sample mean at least this much higher after the training course if there were, in fact, no difference in the population. To make this more informative, it would have been helpful to have the sample means or the sample mean difference and its standard error – or better yet, the confidence interval.

10.97 Which tire is better?

a) There is a range of possible answers for this problem. The standard deviation must be small enough that the resulting standard error and test statistic will lead to a test statistic with a small P-value.

b) The design of this study could be improved by increasing the sample size and randomizing the side the tire is put on.

10.98 Effect of alcoholic parents

a) The groups are dependent since they were matched according to age and gender.

b) 1) Assumptions: the differences in scores are a random sample from a population that is approximately normal.

2) $H_0\colon \mu_d = 0$; $H_a\colon \mu_d \neq 0$

3) $t = \dfrac{2.7}{9.7/\sqrt{49}} = 1.95$

4) The P-value is 0.057.

5) If the null hypothesis is true, the probability of obtaining a difference in sample means at least as extreme as that observed is 0.057. This is some, but not strong, evidence that there is a difference in the mean scores between children of alcoholics versus children of non-alcoholics.

c) We assume that the population of differences is approximately normal and that our sample is a random sample from this distribution.

10.99 CI versus test

a) The 95% confidence interval is $\bar{x}_d \pm t_{.025}(se) = 2.7 \pm 2.01 \pm \left(9.7/\sqrt{49}\right)$, or (–0.1, 5.5), $df = 48$.

We are 95% confident that the difference in population mean scores is between –0.1 and 5.5.

b) The confidence interval gives us a range of values for the difference between the population mean scores rather than just telling us whether or not the scores are significantly different.

10.100 Breast augmentation and self esteem

a) The samples were dependent since the same women were sampled before and after their surgeries.

b) No. In order to find the *t* statistic, we need to know the standard deviation of the differences which cannot be obtained from the information given.

10.101 Internet use

1) Assumptions: the differences in time spent reading news stories on the Internet and time spent communicating on the internet are a random sample from a population that is approximately normal.

2) $H_0\colon \mu_d = 0$; $H_a\colon \mu_d \neq 0$

3) $t = -3.30$

4) The P-value is 0.03.

5) If the null hypothesis is true, the probability of obtaining a difference in sample means at least as extreme as that observed is 0.03. At a significance level of 0.05, we would reject the null hypothesis and conclude that there is a significant difference in the population mean amount of time spent reading news stories on the Internet versus population mean time spent communicating on the Internet.

A 95% confidence interval is given by (–5.9, –0.5). We are 95% confident that the population mean amount of time spent reading news stories on the Internet is between 5.9 and 0.5 hours less than the population mean amount of time spent communicating on the Internet. Since 0 is not contained in this interval, we conclude that the population means differ. The population consists of the 165 students in the course.

10.102 TV or rock music a worse influence?

a) The samples are dependent; the same people are answering both questions.

b) From technology:

The 95% confidence interval is (–0.335, 1.335). We can be 95% confident that the population mean difference between ratings of the influence of TV and rock music is between –0.3 and 1.3. Because 0 falls in this range, it is plausible that there is no difference between the population mean ratings.

10.102 (continued)

c) From technology:

$H_0: \mu_d = 0$; $H_a: \mu_d \neq 0$, $t = 1.32$, and the P-Value is 0.21. The P-value of 0.21 indicates that if the null hypothesis were true, the probability would be 0.21 of getting a test statistic at least as extreme as the value observed.

10.103 Influence of TV and movies

a) We can be 95% confident that the population mean difference between responses with respect to movies and TV is between –0.43 and 0.76. Because 0 falls in this interval, it is plausible that there is no difference between the population mean responses for TV and the population mean responses for movies.

b) The significance test has a P-value of 0.55. If the null hypothesis were true, the probability would be 0.55 of getting a test statistic at least as extreme as the value observed. it is plausible that the null hypothesis is correct and that there is no population mean difference between responses with respect to movies and TV.

10.104 Crossover study

a) The sample proportions are: High Dose: $(53+16)/(53+16+8+9) = 69/86 = 0.802$; Low Dose: $(53+8)/(53+16+8+9) = 61/68 = 0.709$.

b) $z = \dfrac{8-16}{\sqrt{8+16}} = -1.63$; P-value: 0.10; if the null hypothesis were true, the probability would be 0.10 of getting a test statistic at least as extreme as the value observed. It is plausible that the null hypothesis is correct and that there is no difference between low-dose and high-dose analgesics with respect to population proportions who report relief of menstrual bleeding.

The sum of the counts in the denominator should be at least 30, although in practice the two-sided test works well even if this is not true. In addition, the sample should be an independent random sample, and the data should be categorical.

10.105 Belief in ghosts and in astrology

a) Because the same sample is used for both sets of responses, it is not valid to compare the proportions using inferential methods for independent samples. Rather, we should use methods for dependent samples.

b) We do not have enough information to compare the proportions using methods for dependent samples. We would need to know the specific numbers of subjects who said that they did believe in ghosts but not in astrology, and who said that they did believe in astrology, but not in ghosts.

10.106 Death penalty paradox

a) When we ignore victim's race, we observe a proportion of 53/(53 + 414 + 16) = 0.11 white defendants who receive the death penalty, and a proportion of (11 + 4)/(11 + 4 + 37 + 139) = 0.079 (rounds to 0.08) black defendants who receive the death penalty. It appears that whites are more likely to receive the death penalty than are blacks.

When we take victim's race into account, the direction of the association changes with black defendants more likely to get the death penalty.

Specifically, when the victim was white: White defendants have a probability of 53/(53 + 414) = 0.113 (rounds to 0.11) and blacks have a probability of 11/(11 + 37) = 0.229 (rounds to 0.23) of receiving the death penalty.

When the victim was black: White defendants have a probability of 0/(0 + 16) = 0.000 (rounds to 0.00) and blacks have a probability of 4/(4 + 139) = 0.028 (rounds to 0.03) of receiving the death penalty.

In both cases, the proportion of blacks who receive the death penalty is higher than the proportion of whites who receive the death penalty.

b) The death penalty was imposed more frequently when the victim was white, and white victims where more common when the defendant was white.

10.107 Death rate paradoxes

a) If the U.S. has a higher proportion of older people (who are more likely to die in both countries than are younger people), and Mexico has a higher proportion of younger people (who are less likely to die in both countries than are older people), then the death rate in the U.S. could be higher.

b) This could occur if Maine has more older people. Even if older people in South Carolina are more likely to die than are older people in Maine, if Maine has far more older people and far fewer younger people, this overrepresentation of older people could lead to an overall higher death rate in Maine than in South Carolina.

10.108 Income and gender

a) The mean income difference could disappear if we controlled for number of years since receiving highest degree. If most of the female faculty had been hired recently, they would be fewer years from their degree, and would have lower incomes. So, the overall mean could be lower for females. If we look only at those who are a given year from receiving their degree (e.g., received degree five years ago), we might find no gender difference.

b) The mean income difference could disappear if we controlled for college of employment. If more women seek positions in low salary colleges and more men in high salary colleges, it might appear that men make more. If we look only within a given college (e.g., law school), we might not find a gender difference in income.

Chapter Problems: Concepts and Investigations

10.109 Student survey

Each report will be different, but will present findings such as those in the following software output. The assumptions are that each sample must have at least 10 outcomes of each type, the data must be categorical, and the samples must be independent random samples. Using software, we compared the proportions of women and men who said yes. The data, from technology, follow:

From MINITAB:

```
Gender     Y     N    Sample p
f             18    31    0.580645
m           13    29    0.448276
Difference = p (f) - p (m)
Estimate for difference:  0.132369
95% CI for difference:  (-0.118500, 0.383238)
Test for difference = 0 (vs.  not = 0):  Z = 1.03  P-Value =
0.301
```

Since the confidence interval contains 0 and the P-value is large, we are unable to conclude that there is a difference in the population proportion of males and females who believe in life after death.

10.110 Review the medical literature

Answers will vary.

10.111 Attractiveness and getting dates

The short report will be different for each student, but should include the results such as those in the following technology outputs.

Men, From MINITAB:

```
Sample     N    Mean    StDev    SE Mean
1             35      9.7       10.0        1.7
2             36      9.9       12.6        2.1
Difference = mu (1) - mu (2)
Estimate for difference:  -0.200000
95% CI for difference:  (-5.582266, 5.182266)
T-Test of difference = 0 (vs.  not =): T-Value = -0.07  P-Value =
0.941  DF = 66
```

Since the confidence interval contains 0 and the P-value is greater than 0.05, we are unable to conclude that there is a difference in the population mean number of dates between more and less attractive men.

10.111 (continued)

Women, From MINITAB:

Sample	N	Mean	StDev	SE Mean
1	33	17.8	14.2	2.5
2	27	10.4	16.6	3.2

Difference = mu (1) - mu (2)
Estimate for difference: 7.40000
95% CI for difference: (-0.70930, 15.50930)
T-Test of difference = 0 (vs. not =): T-Value = 1.83 P-Value = 0.073 DF = 51

Since the confidence interval contains 0 and the P-value is greater than 0.05, we are unable to conclude that there is a difference in the population mean number of dates between more and less attractive women. However, the confidence interval shows there could be a large difference.

10.112 Pay discrimination against women?

a) We would need to know the sample standard deviations and sample sizes for the two groups.
b) It would not be relevant to conduct a significance test. A significance test lets us make inferences about a population based on a sample. If we already have the information on the entire population, there's no need to make inferences about the population.

10.113 Mean of permutation distribution

If the population distributions are identical, the sampling distribution of $\bar{x}_1 - \bar{x}_2$ should be centered at around 0. The mean is:

$$\sum xP(x) = (-4.67)\left(\frac{1}{10}\right) + (-3.83)\left(\frac{2}{10}\right) + (0.33)\left(\frac{2}{10}\right) + (1.17)\left(\frac{3}{10}\right) + (2.00)\left(\frac{1}{10}\right) + (6.17)\left(\frac{1}{10}\right) = 0.$$

10.114 Treating math anxiety

Solutions will vary but should include the information that follows from technology.
From MINITAB:

```
Two-sample T for Program A vs.  Program B
                   N   Mean  StDev  SE Mean
Program A  5     4.00     2.45      1.1
Program B  5    12.00     3.74      1.7
Difference = mu (Program A) - mu (Program B)
Estimate for difference:  -8.00000
95% CI for difference:  (-12.89382, -3.10618)
T-Test of difference = 0 (vs.  not =): T-Value = -4.00  P-Value =
0.007  DF = 6
```

Since the P-value is quite small (or, equivalently, since the confidence interval are less than 0), we reject the null hypothesis and conclude that there is a significant difference in the drop in the mean number of items that caused anxiety between Programs A and B.

10.115 Obesity and earnings

a) Whether obese (yes or no) and wage are stated to have an association.
b) Education level is one possibility. The women could be paired according to education level and then compared in obesity rates.

10.116 Multiple choice: Alcoholism and gender

The best answer is (b).

10.117 Multiple choice: Comparing mean incomes

The best answer is (d).

10.118 Multiple choice: Sample size and significance

The best answer is (a).

10.119 True or false? Positive values in CI

False

10.120 True or false? Afford food?

False

10.121 True or false? Control for clinic

False

♦♦10.122 Guessing on a test

a) Denote the proportion of correct responses on the test by $\hat{p}_1$ for Joe and by $\hat{p}_2$ for Jane. The sampling distribution of $\hat{p}_1 - \hat{p}_2$ is approximately normal and has a mean of $\hat{p}_1 - \hat{p}_2 = 0.50 - 0.60 = -0.10$ and a standard error of $se = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = \sqrt{\frac{0.6(1-0.6)}{100} + \frac{0.5(1-0.5)}{100}} = 0.070$. The probability that Joe gets a higher score is the probability that $\hat{p}_1 - \hat{p}_2$ is positive. This equals approximately the probability that a normal random variable having a mean of –0.10 and a standard deviation of 0.070 takes a positive value. The z-score for a sample difference of 0 is $z = \frac{0-(-0.10)}{0.070} = \frac{0.10}{0.070} = 1.43$, and the tail probability above that value (which corresponds to positive values for the difference of sample proportions) is 0.08. This is the answer.

b) If the test had only 50 questions, we would have a larger standard error, and hence a smaller z-score and a larger right-tail probability. The probability would be higher than 0.08 that Joe would get a higher score than Jane. (With larger samples, it is less likely for Joe to "luck out" and do better than Jane, even though he has a lower chance of a correct response with any given question.).

♦♦10.123 Standard error of difference

$se(\text{estimate } 1 - \text{estimate } 2) = \sqrt{[se(\text{estimate } 1)]^2 + [se(\text{estimate } 2)]^2} = \sqrt{0.6^2 + 1.8^2} = 1.897$; The 95% confidence interval is $(\bar{x}_1 - \bar{x}_2) \pm 1.96(se) = (46.3 - 33.0) \pm 1.96(1.897)$, or (9.58, 17.02). We can be 95% confident that the difference between the population mean number of years lost is between 9.6 and 17.0. Because 0 is not in this range, we can conclude that there is a population mean difference between those who smoke and are overweight and those who do not smoke and are normal weight in terms of number of years of life left. It appears that those who do not smoke and are normal weight have more years left than do those who smoke and are overweight.

♦♦10.124 Gap between rich and poor: $\sqrt{2/n}$ margin

a) The margin of error is $z_{.025}(se)$, or approximately 2(se), $2(se) = 2\sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}$. With $n_1 = n_2 = n$ and $\hat{p} = 0.5$, $\sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)} = \sqrt{0.5(1-0.5)(1/n + 1/n)} = \sqrt{0.5(0.5)(2/n)} = \sqrt{0.5/n}$. The margin of error is therefore $2\sqrt{0.5/n} = \sqrt{4}\sqrt{0.5/n} = \sqrt{(4)(0.5)/n} = \sqrt{2/n}$.

b) $\sqrt{2/n} = \sqrt{2/1000} = 0.045$. The following pairs of countries are less than 4.5% different from each other: The United States and the United Kingdom, and Turkey and Argentina. The population percentages for these two pairs of countries might not be different.

♦♦10.125 Small-sample CI

a) (i) $\hat{p}_1 = \hat{p}_2 = 0$, because there are no successes in either group (i.e., 0/10 = 0).

(ii) $se = 0$, because there is no variability in either group if all responses are the same. Specifically, both numerators under the square root sign in the se formula would have a zero in them, leading to a calculation of 0 as the se.

(iii) The 95% confidence interval would be (0, 0) because we'd be adding 0 to 0 (se multiplied by z would always be 0 with se of 0, regardless of the confidence level).

10.125 (continued)

b) Using the small-sample method:

(i) $\hat{p}_1 = \hat{p}_2 = 1/12 = 0.083$

(ii) $se = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = \sqrt{\frac{0.083(1-0.083)}{12} + \frac{0.083(1-0.083)}{12}} = 0.113$

The 95% confidence interval is now $(\hat{p}_1 - \hat{p}_2) \pm z_{.025}(se) = (0.083 - 0.083) \pm 1.96(0.113)$, or (–0.22, 0.22), which is far more plausible than (0, 0).

♦♦10.126 Symmetry of permutation distribution

For each permutation that leads to a given difference, you get the same difference but with the opposite sign when all dogs in Group 1 are switched to Group 2 and vice versa. Hence, for each permutation that leads to a positive difference, there is one that gives the same difference (by just switching all observations between groups), but with a negative sign, resulting in a symmetric distribution.

10.127 Null standard error for matched pairs

a) From Table 10.18, $b = 162$, $c = 9$, and $n = 1314$, so the null standard error is $se_0 = \sqrt{(b+c)/n^2} = \sqrt{(162+9)/1314^2} = 0.010$.

b) $z = \frac{(\hat{p}_1 - \hat{p}_1) - 0}{se_0} = \frac{(1117/1314 - 964/1314)}{0.010} = 11.7 = \frac{162-9}{\sqrt{162+9}} = \frac{b-c}{\sqrt{b+c}}$

♦♦10.128 Graphing Simpson's paradox

a) A comparison of a pair of circles having the same letter in the middle (e.g., W) indicates that the death penalty was more likely for black than white defendants, when we control for victim's race. For both cases with white victims and with black victims, a higher percentage of black defendants received the death penalty; the circles for black defendants with victims of a given race are higher on the *y*-axis than the circles for white defendants with victims of that same race.

b) When we compare the **x** marks for black and white defendants without regard to victim's race, the mark for white defendants is higher with respect to the *y*-axis than is the mark for black defendants.

c) For white defendants, the overall percentage who got the death penalty is so close to the percentage for the case with white victims because almost all white defendants are accused of killing white victims.

Chapter Problems: Student Activities

10.129 Reading the medical literature

The reports will differ based on the article chosen by the class instructor.