Answers to Chapter 10 Review Exercises

R10.1 (a) Paired t test for the mean difference. (b) Two-sample z interval for the difference in proportions. (c) One-sample t interval for the mean. (d) Two-sample t interval for the difference between two means.

R10.2 (a) Observational study; drivers were not assigned to the cities. (b) **State:** Our parameters are p_1 and p_2, the actual proportions of Hispanic female drivers in New York and in Boston, respectively, who wear seat belts. We want to estimate $p_1 - p_2$ at a 95% confidence level. **Plan:** Two-sample z interval for $p_1 - p_2$. *Random:* The women were selected randomly. *Normal:* 183, 37, 68, 49 are all at least 10. *Independent:* There are more than 2200 Hispanic women drivers in New York and 1170 Hispanic women drivers in Boston. **Do:** $(0.832 - 0.581) \pm 1.96 \sqrt{\frac{0.832(0.168)}{220} + \frac{0.581(0.419)}{117}} = (0.149, 0.353)$. **Conclude:** We are 95% confident that the interval from 0.149 to 0.353 captures the difference in actual proportions of Hispanic women drivers in New York and Boston who wear their seat belts. (c) Since 0 is not in the interval, we have good evidence that a smaller proportion of Hispanic women in Boston wear their seat belts.

R10.3 (a) The Random and Independent conditions are met because this is a randomized comparative experiment. (b) The sample sizes are large enough: $n_1 = n_2 = 45$. (c) If there were no difference in the ratings of the product under the two treatments, we would have less than a 1% chance of observing a difference as large as or larger than the one in this experiment.

R10.4 **State:** $H_0: \mu_1 - \mu_2 = 0$ versus $H_a: \mu_1 - \mu_2 \neq 0$, where μ_1 and μ_2 are the actual mean NAEP quantitative skills test scores for men and for women, respectively. **Plan:** Two-sample t test. *Random:* Both samples were randomly selected. *Normal:* Both samples had at least 30 observations. *Independent:* There are more than 8400 males who took the NAEP and 10,770 women who took the NAEP. **Do:** From the data, $n_1 = 840$, $\bar{x}_1 = 272.40$, $s_1 = 59.2$, $n_2 = 1077$, $\bar{x}_2 = 274.73$, and $s_2 = 57.5$. Using df $= 839$: $t = -0.865$ and P-value $= 2P(t < -0.865) = 0.3872$. **Conclude:** Since P-value > 0.01, we fail to reject H_0. We do not have enough evidence to conclude that males and females have different mean scores on the NAEP quantitative skills test.

R10.5 (a) **State:** $H_0: p_1 - p_2 = 0$ versus $H_a: p_1 - p_2 < 0$, where p_1 and p_2 are the actual proportions of patients like the ones in this study who develop AIDS after taking AZT and after taking a placebo, respectively. **Plan:** Two-sample z test for $p_1 - p_2$. *Random:* This was a randomized comparative experiment. *Normal:* 17, 418, 38, 397 are all at least 10. *Independent:* Due to the random assignment, these two groups of patients can be viewed as independent. Also, knowing whether one patient developed AIDS gives no information about whether another patient developed AIDS. **Do:** From the data, $\hat{p}_1 = 0.039$ and $\hat{p}_2 = 0.087$. $\hat{p}_C = 0.063$, $z = -2.91$, P-value $= P(z < -2.91) = 0.0018$. **Conclude:** Since P-value < 0.05, we reject H_0. We have enough evidence to conclude that taking AZT lowers the proportion of patients like these who develop AIDS. (b) Neither the subjects nor the researchers who had contact with them (including those determining whether subjects had AIDS) knew which subjects were getting which drug. This is important because if the experiment were not double-blind, the results could not be attributed to the treatments. (c) A Type I error means concluding that AZT lowers the risk of developing AIDS when in fact it does not. A Type II error means concluding that we do not have enough evidence that AZT lowers the risk of developing AIDS when in fact it does. Answers may vary as to which one is more serious.

R10.6 (a) Normal condition is not met. (b) Normal condition is not met.

R10.7 (a) The students in the treatment group had generally higher differences in scores, and the middle 50% of their differences was much more compact than for the control group. The treatment group differences were also reasonably symmetric, whereas the control group differences were more right-skewed. (b) **State:** Our parameters are μ_1 and μ_2, the actual mean difference in test scores for students like these who get the treatment message and who get the neutral message, respectively. We want to estimate $\mu_1 - \mu_2$ at a 90% confidence level. **Plan:** Two-sample t interval for $\mu_1 - \mu_2$. *Random:* This was a randomized controlled experiment. *Normal:* Neither boxplot shows strong skewness or outliers. *Independent:* Due to the random assignment, these two groups of students can be viewed as independent. Also, knowing one student's difference in scores gives no information about another student's difference in scores. **Do:** Using (post − pre) as our difference, $n_1 = 10$, $\bar{x}_1 = 11.4$, $s_1 = 3.169$, $n_2 = 8$, $\bar{x}_2 = 8.25$, and $s_2 = 3.69$. Using df $= 7$: $(11.4 - 8.25) \pm 1.895 \sqrt{\frac{(3.169)^2}{10} + \frac{(3.69)^2}{8}} = (0.03, 6.27)$. **Conclude:** We are 90% confident that the interval from 0.03 to 6.27 captures the true difference in mean scores for those who received the treatment message and those who received the neutral message. (c) We cannot generalize to all students who failed the test, because our sample was not a random sample of all students who failed the test. It was a group of students who agreed to participate in the experiment.

R10.8 (a) About 139 (b) Properties of the sampling distribution of the difference $\hat{p}_1 - \hat{p}_2$ can be obtained from properties of the individual sampling distributions used in the individual confidence intervals, but the upper and lower limits of the intervals are not directly related. (c) **State:** Our parameters are p_1 and p_2, the actual proportions of golf clubs returned in 5 days or fewer before the changes and after the changes, respectively. We want to estimate $p_1 - p_2$ at a 95% confidence level. **Plan:** Two-sample z interval for $p_1 - p_2$. *Random:* The clubs both before and after changes were selected randomly. *Normal:* 22, 117, 125, 14 are all at least 10. *Independent:* There are more than 1390 golf clubs repaired before the changes and 1390 golf clubs repaired after the changes. **Do:** From the data, $n_1 = 139$, $\hat{p}_1 = 0.16$, $n_2 = 139$, and $\hat{p}_2 = 0.90$. $(0.90 - 0.16) \pm$

$$1.96\sqrt{\frac{0.90(0.10)}{139} + \frac{0.16(0.84)}{139}} =$$

$(0.661, 0.819)$ **Conclude:** We are 95% confident that the interval from 0.661 to 0.819 captures the difference in the actual proportions of orders sent back to customers in 5 days or fewer before the changes and after the changes.

R10.9 (a) Experiment; a treatment (message or no message) was deliberately imposed on the subjects. (b) **State:** We want to perform a test at the $\alpha = 0.05$ significance level of $H_0: p_1 - p_2 = 0$ versus $H_a: p_1 - p_2 > 0$, where p_1 is the actual proportion of people who would be contacted when a message is left and p_2 is the actual proportion of people who would be contacted when a message is not left.

Plan: Use a two-sample z test for $p_1 - p_2$ if the conditions are satisfied. *Random:* This was a randomized comparative experiment. *Normal:* The numbers of successes and failures in both groups are at least 10 (Message: 200 successes, 91 failures; No message: 58 successes, 42 failures). *Independent:* Due to the random assignment, these two groups of people contacted can be viewed as independent. Also, knowing whether one person responded gives no information about whether another person responded. **Do:** From the data, $\hat{p}_1 = 0.687$ and $\hat{p}_2 = 0.58$. The pooled proportion is $\hat{p}_C = 0.66$. The test statistic is $z = 1.95$, and the P-value is $P(z > 1.95) = 0.0256$. **Conclude:** Since P-value < 0.05, we reject H_0. We have enough evidence to conclude that the proportion of people like the ones in the study who would respond when messages are left is higher than the proportion of people who would respond when no messages are left.

R10.10 Answers will vary, but here is an example. The difference between average female (55.5) and male (57.9) self-concept scores was so small that it can be attributed to chance variation in the samples ($t = -0.83$, df $= 62.8$, P-value $= 0.4110$). In other words, based on this sample, we have insufficient evidence to suggest that mean self-concept scores differ by gender.

Chapter 10 AP Statistics Practice Test

T10.1 e

T10.2 b

T10.3 a

T10.4 a

T10.5 b

T10.6 e

T10.7 c

T10.8 c

T10.9 b

T10.10 a

T10.11 (a) **State:** Our parameters of interest are μ_1 = the mean hospital stay for patients like those who get heating blankets during surgery and μ_2 = the mean hospital stay for patients like those who have core temperatures reduced during surgery. We want to estimate $\mu_1 - \mu_2$ at a 95% confidence level. **Plan:** Use a two-sample t interval for $\mu_1 - \mu_2$ if the conditions are satisfied. *Random:* This was a randomized controlled experiment. *Normal:* Both n_1 and n_2 are at least 30. *Independent:* Due to the random assignment, these two groups of patients can be viewed as independent. Also, knowing one patient's length of hospital stay gives no information about another patient's length of hospital stay. **Do:** From the data, $n_1 = 104$, $\bar{x}_1 = 12.1$, $s_1 = 4.4$, $n_2 = 96$, $\bar{x}_2 = 14.7$, and $s_2 = 6.5$. Using the conservative df $= 95$, the 95% confidence interval is $(12.1 - 14.7) \pm 1.985$

$$\sqrt{\frac{(4.4)^2}{104} + \frac{(6.5)^2}{96}} = (-4.17, -1.03).$$

Conclude: We are 95% confident that the interval from -4.17 to -1.03 captures the difference in actual mean hospital stay for patients like those who get heating blankets during surgery and those who have their core temperatures reduced during surgery. This interval suggests that the mean hospital stay is between 4.17 and 1.03 days shorter for those receiving the heating blankets than for those who have their core temperatures reduced. (b) Yes. Since 0 is not in the interval, the entire interval is negative and we subtracted normothermic − hypothermic. It appears that warming patients during surgery decreases the mean hospital stay.

T10.12 (a) **State:** We want to perform a test at the $\alpha = 0.05$ significance level of $H_0: p_1 - p_2 = 0$ versus $H_a: p_1 - p_2 > 0$, where p_1 is the actual proportion of cars that had the brake defect in last year's model and p_2 is the actual proportion of cars that have the brake defect in this year's model. **Plan:** Use a two-sample z test for $p_1 - p_2$ if the conditions are satisfied. *Random:* The samples were randomly selected. *Normal:* The number of successes and failures in both groups are at least 10 (last year: 20 successes, 80 failures; this year: 50 successes, 300 failures). *Independent:* Both samples are less than 10% of their respective populations (there are more than 1000 cars of last year's model and 3500 cars of this year's model). **Do:** From the data, $\hat{p}_1 = 0.2$ and $\hat{p}_2 = 0.143$. The pooled proportion is $\hat{p}_C = 0.156$. The test statistic is $z = 1.39$, and the P-value is $P(z > 1.39) = 0.0823$.

Conclude: Since P-value > 0.05, we fail to reject H_0. We do not have enough evidence to conclude that the proportion of brake defects is less in cars of this year's model than in cars of last year's model. (b) A Type I error occurs if we reject the null hypothesis and it is really true. In this case, that would mean concluding that there are fewer brake defects in this year's car model when in fact there are not fewer. This might result in more accidents because people think that their brakes are safe. A Type II error occurs if we fail to reject the null hypothesis when in fact it is false. In this case, that would mean concluding that there are no fewer brake defects this year than last year when the number of brake defects has actually been reduced. This might result in