Sections 13.1 - 13.2

1. Which statement is not true about hypothesis tests?
   A. Hypothesis tests are only valid when the sample is representative of the population for the question of interest.
   B. Hypotheses are statements about the population represented by the samples.
   C. Hypotheses are statements about the sample (or samples) from the population.
   D. Conclusions are statements about the population represented by the samples.
   KEY: C

2. The primary purpose of a significance test is to
   A. estimate the p-value of a sample.
   B. estimate the p-value of a population.
   C. decide whether there is enough evidence to support a research hypothesis about a sample.
   D. decide whether there is enough evidence to support a research hypothesis about a population.
   KEY: D

3. The level of significance associated with a significance test is the probability
   A. of rejecting a true null hypothesis.
   B. of not rejecting a true null hypothesis.
   C. that the null hypothesis is true.
   D. that the alternative hypothesis is true.
   KEY: A

4. A result is called statistically significant whenever
   A. the null hypothesis is true.
   B. the alternative hypothesis is true.
   C. the p-value is less or equal to the significance level.
   D. the p-value is larger than the significance level.
   KEY: C

5. Which of the following is not one of the steps for hypothesis testing?
   A. Determine the null and alternative hypotheses.
   B. Verify data conditions and calculate a test statistic.
   C. Assuming the null hypothesis is true, find the p-value.
   D. Assuming the alternative hypothesis is true, find the p-value.
   KEY: D

6. Which of the following is not a correct way to state a null hypothesis?
   A. $H_0: \mu_c$
   B. $H_0: \mu = 10$
   C. $H_0: \mu = 0$
   D. $H_0: \mu = 0.5$
   KEY: A
7. In hypothesis testing for one mean, the "null value" is not used in which of the following?
   A. The null hypothesis.
   B. The alternative hypothesis.
   C. The computation of the test statistic.
   D. The (null) standard error.
   KEY: D

8. A null hypothesis is that the average pulse rate of adults is 70. For a sample of 64 adults, the average pulse rate is 71.8. A significance test is done and the p-value is 0.02. What is the most appropriate conclusion?
   A. Conclude that the population average is 70.
   B. Conclude that the population average is 71.8.
   C. Reject the hypothesis that the population average is 70.
   D. Reject the hypothesis that the sample average is 70.
   KEY: C

9. The p-value for a one-sided test for a mean was 0.04. The p-value for the corresponding two-sided test would be:
   A. 0.02
   B. 0.04
   C. 0.06
   D. 0.08
   KEY: D

10. A null hypothesis is that the mean cholesterol level is 200 in a certain age group. The alternative is that the mean is not 200. Which of the following is the most significant evidence against the null and in favor of the alternative?
    A. For a sample of \( n = 25 \), the sample mean is 220.
    B. For a sample of \( n = 10 \), the sample mean is 220.
    C. For a sample of \( n = 50 \), the sample mean is 180.
    D. For a sample of \( n = 20 \), the sample mean is 180.
    KEY: C

11. A test of \( H_0: \mu = 0 \) versus \( H_a: \mu > 0 \) is conducted on the same population independently by two different researchers. They both use the same sample size and the same value of \( \alpha = 0.05 \). Which of the following will be the same for both researchers?
    A. The p-value of the test.
    B. The power of the test if the true \( \mu = 6 \).
    C. The value of the test statistic.
    D. The decision about whether or not to reject the null hypothesis.
    KEY: B

12. Which of the following is not true about hypothesis testing?
    A. The null hypothesis defines a specific value of a population parameter, called the null value.
    B. A relevant statistic is calculated from population information and summarized into a “test statistic.”
    C. A p-value is computed on the basis of the standardized “test statistic.”
    D. On the basis of the p-value, we either reject or fail to reject the null hypothesis.
    KEY: B
Questions 13 to 17: An investigator wants to assess whether the mean $\mu$ = the average weight of passengers flying on small planes exceeds the FAA guideline of average total weight of 185 pounds (passenger weight including shoes, clothes, and carry-on). Suppose that a random sample of 51 passengers showed an average total weight of 200 pounds with a sample standard deviation of 59.5 pounds. Assume that passenger total weights are normally distributed.

13. What are the appropriate null and alternative hypotheses?
   A. $H_0: \mu = 185$ and $H_a: \mu < 185$
   B. $H_0: \mu = 185$ and $H_a: \mu > 185$.
   C. $H_0: \mu = 185$ and $H_a: \mu \neq 185$.
   D. $H_0: \mu \neq 185$ and $H_a: \mu = 185$.
   KEY: B

14. What is the value of the test statistic?
   A. $t = 1.50$
   B. $t = 1.65$
   C. $t = 1.80$
   D. None of the above
   KEY: C

15. What is the $p$-value?
   A. $p$-value = 0.039
   B. $p$-value = 0.053
   C. $p$-value = 0.070
   D. None of the above
   KEY: A

16. For a significance level of $\alpha = 0.05$, are the results statistically significant?
   A. No, the results are not statistically significant because the $p$-value < 0.05.
   B. Yes, the results are statistically significant because the $p$-value < 0.05.
   C. No, the results are not statistically significant because the $p$-value > 0.05
   D. Yes, the results are statistically significant because the $p$-value > 0.05.
   KEY: B

17. Which of the following is an appropriate conclusion?
   A. The results are statistically significant so the average total weight of all passengers appears to be greater than 185 pounds.
   B. The results are statistically significant so the average total weight of all passengers appears to be less than 185 pounds.
   C. The results are not statistically significant so there is not enough evidence to conclude the average total weight of all passengers is greater than 185 pounds.
   D. None of the above.
   KEY: A
18. A random sample of 25 third graders scored an average of 3.2 on a standardized reading test. The standard deviation was 0.95. What is the value of the \( t \)-test statistic for determining if the mean score is significantly higher than 3?
   A. \( t = 0.21 \)
   B. \( t = 5.26 \)
   C. \( t = 1.05 \)
   D. None of the above
   **KEY: C**

19. The WISC scores of a sample of \( n = 20 \) 5th graders resulted in a mean of 104 with a standard deviation of 9. Is the mean score significantly higher than 100 when \( \alpha = 0.05 \)?
   A. No, because \( t = 1.99 \) which is greater than the critical value \( t^* = 1.73 \).
   B. Yes, because \( t = 1.99 \) which is greater than the critical value \( t^* = 1.73 \).
   C. No, because \( t = 1.99 \) which is smaller than the critical value \( t^* = 2.09 \).
   D. Yes, because \( t = 1.99 \) which is smaller than the critical value \( t^* = 2.09 \).
   **KEY: B**

20. A random sample of 25 third graders scored an average of 3.3 on a standardized spelling test. The standard deviation was 1.06. Is the mean score significantly higher than 3?
   A. It is both when \( \alpha = 0.05 \) and when \( \alpha = 0.10 \).
   B. It is when \( \alpha = 0.05 \) but it is not when \( \alpha = 0.10 \).
   C. It is not when \( \alpha = 0.05 \) but it is when \( \alpha = 0.10 \).
   D. It is not when \( \alpha = 0.05 \) and it is not either when \( \alpha = 0.10 \).
   **KEY: C**

21. A random sample of 12 bags of barbecue potato chips is collected and the price of each determined. The mean was $1.03 with a standard deviation of 7 cent ($0.07). Using \( \alpha = 0.10 \), we wish to test \( H_0: \mu = $1.00 \). The result is significant
   A. for both \( H_a: \mu > $1.00 \) and for \( H_a: \mu \neq $1.00 \).
   B. for \( H_a: \mu > $1.00 \) but not for \( H_a: \mu \neq $1.00 \).
   C. for \( H_a: \mu \neq $1.00 \) but not for \( H_a: \mu > $1.00 \).
   D. neither for \( H_a: \mu > $1.00 \) nor for \( H_a: \mu \neq $1.00 \).
   **KEY: B**

22. Spatial perception is measured on a scale from 0 to 10. A group of 9th grade students are tested for spatial perception. SPSS was used to obtain descriptive statistics of the spatial perception scores in the sample.

<table>
<thead>
<tr>
<th>One-Sample Statistics</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial Perception</td>
<td>18</td>
<td>6.00</td>
<td>1.815</td>
<td>.428</td>
</tr>
</tbody>
</table>

Using \( \alpha = 0.01 \), is the mean score significantly different from 5?
   A. Yes, because \( t = 2.34 \) and this is greater than \( t^* = 2.11 \).
   B. No, because \( t = 2.34 \) and this is smaller than \( t^* = 3.97 \).
   C. No, because \( t = 0.55 \) and this is smaller than \( t^* = 2.11 \).
   D. No, because \( t = 2.34 \) and this is smaller than \( t^* = 2.90 \).
   **KEY: D**
Questions 23 to 27: A safety officer wants to prove that \( \mu \) = the average speed of cars driven by a school is less than 25 mph. Suppose that a random sample of 14 cars shows an average speed of 24.0 mph, with a sample standard deviation of 2.2 mph. Assume that the speeds of cars are normally distributed.

23. What are the appropriate null and alternative hypotheses?
   A. \( H_0: \mu = 25 \) and \( H_a: \mu < 25 \)
   B. \( H_0: \mu = 25 \) and \( H_a: \mu > 25 \)
   C. \( H_0: \mu = 25 \) and \( H_a: \mu \neq 25 \)
   D. \( H_0: \mu 
eq 25 \) and \( H_a: \mu = 25 \)
   KEY: A

24. What is the value of the test statistic?
   A. \( t = -1.80 \)
   B. \( t = -1.70 \)
   C. \( t = -1.50 \)
   D. None of the above
   KEY: B

25. What is the \( p \)-value?
   A. \( p \)-value = 0.079
   B. \( p \)-value = 0.057
   C. \( p \)-value = 0.048
   D. None of the above
   KEY: B

26. For a significance level of \( \alpha = 0.05 \), are the results statistically significant?
   A. No, the results are not statistically significant because the \( p \)-value < 0.05.
   B. Yes, the results are statistically significant because the \( p \)-value < 0.05.
   C. No, the results are not statistically significant because the \( p \)-value > 0.05
   D. Yes, the results are statistically significant because the \( p \)-value > 0.05.
   KEY: C

27. Which of the following is an appropriate conclusion?
   A. The results are statistically significant so the average speed appears to be greater than 25 mph.
   B. The results are statistically significant so the average speed appears to be less than 25 mph.
   C. The results are not statistically significant: there is not enough evidence to conclude the average speed is less than 25 mph.
   D. None of the above.
   KEY: C
Questions 28 to 32: A counselor wants to show that for men who are married by the time they are 30, \( \mu = \) average age when the men are married is not 21 years old. A random sample of 10 men who were married by age 30 showed an average age at marriage of 22.2, with a sample standard deviation of 1.9 years. Assume that the age at which this population of men get married for the first time is normally distributed.

28. What are the appropriate null and alternative hypotheses?
   A. \( H_0: \mu = 21 \) and \( H_a: \mu < 21 \)
   B. \( H_0: \mu = 21 \) and \( H_a: \mu > 21 \)
   C. \( H_0: \mu = 21 \) and \( H_a: \mu \neq 21 \)
   D. \( H_0: \mu \neq 21 \) and \( H_a: \mu = 21 \)
   KEY: C

29. What is the value of the test statistic?
   A. \( t = 1.80 \)
   B. \( t = 2.00 \)
   C. \( t = 2.33 \)
   D. None of the above
   KEY: B

30. What is the \( p \)-value?
   A. \( p \)-value = 0.022
   B. \( p \)-value = 0.053
   C. \( p \)-value = 0.076
   D. None of the above
   KEY: C

31. For a significance level of \( \alpha = 0.05 \), are the results statistically significant?
   A. No, the results are not statistically significant because the \( p \)-value < 0.05.
   B. Yes, the results are statistically significant because the \( p \)-value < 0.05.
   C. No, the results are not statistically significant because the \( p \)-value > 0.05
   D. Yes, the results are statistically significant because the \( p \)-value > 0.05.
   KEY: C

32. Which of the following is an appropriate conclusion?
   A. The results are statistically significant so the average age appears to be greater than 21.
   B. The results are statistically significant so the average age appears to be less than 21.
   C. The results are not statistically significant so there is not enough evidence to conclude average age is different from 21.
   D. None of the above.
   KEY: C
33. Explain the difference between the null and alternative hypotheses in hypothesis testing. Give an example.
KEY: The typical null hypothesis specifies a specific value for a population parameter; the alternative hypothesis specifies a range of values. The null hypothesis describes the status quo, or the condition of no difference in a paired design. The alternative hypothesis describes a change, or the condition of a difference in a paired design. For testing about the value of a population mean \( \mu \), a null hypothesis could be \( H_0: \mu = 10 \) and the alternative hypothesis could be \( H_a: \mu \neq 10 \).

**Questions 34 to 37:** A sample of \( n = 9 \) men are asked, "What's the fastest you've ever driven a car?" The sample mean is 120 mph and the standard deviation is 30.

34. What is the value of the \( t \)-test statistic for testing the null hypothesis that the population mean response is 100 mph?
KEY: The value of the \( t \)-test statistic is 2.00.

35. What are the degrees of freedom for conducting this \( t \)-test?
KEY: The degrees of freedom are \( n - 1 = 9 - 1 = 8 \).

36. Find the \( p \)-value for testing whether the population mean response is more than 100 mph?
KEY: The corresponding \( p \)-value = 0.040.

37. Give an appropriate conclusion using a 1\% level of significance.
KEY: The results are not statistically significant so the null hypothesis cannot be rejected. There is not enough evidence to conclude that the population mean is more than 100 mph.

**Questions 38 to 40:** A group of women suffering from various types of obsessive fear is given a therapeutic course to learn how to handle their fear. At the beginning of the course the women are given a questionnaire to determine how afraid they are. Fear scores range between 100 and 400. SPSS is used to obtain descriptive statistics. Output is shown below.

<table>
<thead>
<tr>
<th>Descriptive Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Obsessive Fear</td>
</tr>
<tr>
<td>Valid N (listwise)</td>
</tr>
</tbody>
</table>

The course coordinator wishes to test if the women score significantly lower than 250.

38. What is the value of the \( t \)-test statistic?
KEY: \( t = -0.214 \)

39. What do we know about the \( p \)-value for this test?
KEY: \( p \)-value > 0.10 (\( p \)-value = 0.416)

40. Give an appropriate conclusion using a 10\% level of significance.
KEY: The results are not statistically significant. There is not enough evidence to conclude that the population mean fear score is lower than 250.
Section 13.3

41. The amount of time the husband and the wife spend on house work is measured for 15 women and their 15 husbands. For the wives the mean was 7 hours/week and for the husbands the mean was 4.5 hours/week. The standard deviation of the differences in time spent on house work was 2.85. What is the value of the test statistic for testing the difference in mean time spent on housework between husbands and wives?
   A. 0.88
   B. 2.40
   C. 3.40
   D. 4.80
   KEY: C

42. The head circumference is measured for 25 girls and their younger twin sisters. The mean of the older twin girls was 50.23 cm and the mean of the younger twins was 49.96 cm. The standard deviation of the differences was 1 cm. Is this difference significant at a significance level of 5%?
   A. Yes, because $t = 1.35$ and the $p$-value < 0.10.
   B. Yes, because $t = 6.25$ and the $p$-value < 0.10.
   C. No, because $t = 1.35$ and the $p$-value > 0.10.
   D. No, because $t = 0.27$ and the $p$-value > 0.10.
   KEY: C

43. An experiment is conducted with 15 seniors who are taking Spanish at Oak View High School. A randomly selected group of eight students is first tested with a written test and a day later with an oral exam. To avoid order effects, the other seven students are tested in reverse order. The instructor is interested in the difference in grades between the two testing methods. SPSS is used to obtain descriptive statistics for the grades of the two tests.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Talk</td>
<td>15</td>
<td>1.523</td>
<td>1.530</td>
</tr>
<tr>
<td>Write</td>
<td>15</td>
<td>4.166</td>
<td>2.047</td>
</tr>
<tr>
<td>Talk - Write</td>
<td>15</td>
<td>-2.643</td>
<td>2.182</td>
</tr>
</tbody>
</table>

Is there a significant difference between the mean grades using the two different testing methods? Use $\alpha = 0.05$.
   A. Yes, because $t = -4.69$ and the $p$-value < 0.05.
   B. No, because $t = -1.21$ and the $p$-value > 0.05.
   C. Yes, because $t = -6.63$ and the $p$-value < 0.05.
   D. No, because $t = -1.48$ and the $p$-value > 0.05.
   KEY: A
Questions 44 to 47: It is known that for right-handed people, the dominant (right) hand tends to be stronger. For left-handed people who live in a world designed for right-handed people, the same may not be true. To test this, muscle strength was measured on the right and left hands of a random sample of 15 left-handed men and the difference (left - right) was found. The alternative hypothesis is one-sided (left hand stronger). The resulting t-statistic was 1.80.

44. This is an example of
   A. a two-sample t-test.
   B. a paired t-test.
   C. a pooled t-test.
   D. an unpooled t-test.

KEY: B

45. Which of the following is true about the conditions necessary to carry out the t-test in this situation?
   A. Because the sample size is small (15), the population of differences must be assumed to be approximately normal, but no assumption about the variances is required.
   B. Because the sample size is not small (15 + 15 = 30), the population of differences need not be assumed to be approximately normal, and no assumption about the variances is required.
   C. Because the sample size is small (15), the population of differences must be assumed to be approximately normal, and the variances of the right and left hand strengths must be assumed to be equal.
   D. Because the sample size is not small (15 + 15 = 30), the population of differences need not be assumed to be approximately normal, but the variances of the right and left hand strengths must be assumed to be equal.

KEY: A

46. Assuming the conditions are met, based on the t-statistic of 1.80 the appropriate decision for this test using \( \alpha = 0.05 \) is:
   A. df = 14, so p-value < 0.05 and the null hypothesis can be rejected.
   B. df = 14, so p-value > 0.05 and the null hypothesis cannot be rejected.
   C. df = 28, so p-value < 0.05 and the null hypothesis can be rejected.
   D. df = 28, so p-value > 0.05 and the null hypothesis cannot be rejected.

KEY: A

47. Which of the following is an appropriate conclusion?
   A. The results are statistically significant so the left hand appears to be stronger.
   B. The results are statistically significant so the left hand does not appear to be stronger.
   C. The results are not statistically significant so there is not enough evidence to conclude the left hand appears to be stronger.
   D. None of the above.

KEY: A
Questions 48 to 50: A group of women suffering from various types of obsessive fear is given a therapeutic course to learn how to handle their fear. At the end of the course the women are given a questionnaire to determine how afraid they are. Six months after the course the women answer the questionnaire again as a follow-up measure. The fear-scores are typed into SPSS. Output is shown below.

<table>
<thead>
<tr>
<th>Paired Samples Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paired Differences</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Pair 1 post-follow</td>
</tr>
</tbody>
</table>

48. How much lower, on average, did the women score on the fear questionnaire six months after the course than directly at the end of the course?
KEY: $\overline{d} = 2.52$

49. What is the value of the standard deviation of the differences that is missing from the output?
KEY: $s_d = 4.415$

50. What is the $p$-value for testing whether women score significantly lower on fear 6 months after the course than directly at the end of the course?
KEY: $p$-value = 0.0044.

Questions 51 to 54: An instructor wanted to prove that writing with the non-dominant hand takes longer on average than writing with the dominant hand. A significance level of 5% will be used. Ten students randomly sampled from a large class each wrote the word “Statistics” twice; once with their dominant hand and once with their non-dominant hand, in random order. The average time to write using their dominant hands was 5 seconds, while the average time to write with their non-dominant hand was 10 seconds. The standard deviation of the difference was 7.9 seconds, and the standard error was 2.498. Assume that writing times with each hand are normally distributed.

51. What is the value of the $t$-test statistic?
KEY: The test statistic is $t = 2.00$.

52. What are the degrees of freedom for conducting this $t$-test?
KEY: The degrees of freedom are $n - 1 = 10 - 1 = 9$.

53. Find the $p$-value for testing whether using the non-dominant hand takes longer on average than using the dominant hand?
KEY: The corresponding $p$-value = 0.038.

54. Given an appropriate conclusion using a 5% level of significance.
KEY: The results are statistically significant so the non-dominant hand appears to take longer on average than using the dominant hand.
**Questions 55 to 58:** An instructor wanted to prove that writing with the non-dominant hand takes longer on average than writing with the dominant hand. A significance level of 5% will be used. Ten students randomly sampled from a large class each wrote the word “Statistics” twice; once with their dominant hand and once with their non-dominant hand, in random order. The average time to write using their dominant hands was 5 seconds, while the average time to write with their non-dominant hand was 10 seconds. The standard deviation of the difference was 7.9 seconds, and the standard error was 2.498. Assume that writing times with each hand are normally distributed.

55. What is the value of the $t$-test statistic?
**KEY:** The test statistic is $t = 2.00$.

56. What are the degrees of freedom for conducting this $t$-test?
**KEY:** The degrees of freedom are $n - 1 = 10 - 1 = 9$.

57. Find the $p$-value for testing whether using the non-dominant hand takes longer on average than using the dominant hand?
**KEY:** The corresponding $p$-value = 0.038.

58. Given an appropriate conclusion using a 5% level of significance.
**KEY:** The results are statistically significant so the non-dominant hand appears to take longer on average than using the dominant hand.
Section 13.4

59. Suppose that a difference between two groups is examined. In the language of statistics, the *alternative hypothesis* is a statement that there is
   A. no difference between the sample means.
   B. a difference between the sample means.
   C. no difference between the population means.
   D. a difference between the population means.
   KEY: D

60. The maximum distance at which a highway sign can be read is determined for a sample of young people and a sample of older people. The mean distance is computed for each age group. What's the most appropriate null hypothesis about the means of the two groups?
   A. The population means are different.
   B. The sample means are different.
   C. The population means are the same.
   D. The sample means are the same.
   KEY: C

61. Researchers want to see if men have a higher blood pressure than women do. A study is planned in which the blood pressures of 50 men and 50 women will be measured. What's the most appropriate alternative hypothesis about the means of the men and women?
   A. The sample means are the same.
   B. The sample mean will be higher for men.
   C. The population means are the same.
   D. The population mean is higher for men than for women.
   KEY: D

62. When comparing two means, the situation most likely to lead to a result that is statistically significant but of little practical importance is
   A. when the actual difference is large and the sample sizes are large.
   B. when the actual difference is large and the sample sizes are small.
   C. when the actual difference is small and the sample sizes are large.
   D. when the actual difference is small and the sample sizes are small.
   KEY: C

63. When comparing two means, which situation is most likely to lead to a result that is statistically significant?
   (Consider all other factors equal, such as significance level and standard deviations.)
   A. $\bar{x}_1 = 10$, $\bar{x}_2 = 20$ and the sample sizes are $n_1 = 25$ and $n_2 = 25$
   B. $\bar{x}_1 = 10$, $\bar{x}_2 = 25$ and the sample sizes are $n_1 = 25$ and $n_2 = 25$
   C. $\bar{x}_1 = 10$, $\bar{x}_2 = 20$ and the sample sizes are $n_1 = 15$ and $n_2 = 20$
   D. $\bar{x}_1 = 10$, $\bar{x}_2 = 25$ and the sample sizes are $n_1 = 25$ and $n_2 = 35$
   KEY: D
Questions 64 and 65: A null hypothesis is that the mean nose lengths of men and women are the same. The alternative hypothesis is that men have a longer mean nose length than women.

64. Which of the following is the correct way to state the null hypothesis?
   A. \( H_0: p = 0.50 \)
   B. \( H_0: \bar{x}_1 - \bar{x}_2 = 0 \)
   C. \( H_0: \mu_1 - \mu_2 = 0 \)
   D. \( H_0: \mu_1 - \mu_2 = 0 \)

   KEY: D

65. A statistical test is performed for assessing if men have a longer mean nose length than women. The p-value is 0.225. Which of the following is the most appropriate way to state the conclusion?
   A. The mean nose lengths of the populations of men and women are identical.
   B. There is not enough evidence to say that the populations of men and women have different mean nose lengths.
   C. Men have a greater mean nose length than women.
   D. The probability is 0.225 that men and women have the same mean nose length.

   KEY: B

Questions 66 to 70: An airport official wants to assess if the flights from one airline (Airline 1) are less delayed than flights from another airline (Airline 2). Let \( \mu_1 \) = average delay for Airline 1 and \( \mu_2 \) = average delay for Airline 2. A random sample of 10 flights for Airline 1 shows an average of 9.5 minutes delay with a standard deviation of 3 minutes. A random sample of 10 flights for Airline 2 shows an average of 12.63 minutes delay with a standard deviation of 3 minutes. Assume delay times are normally distributed, but do not assume the population variances are equal. Use the conservative “by hand” estimate for the degrees of freedom.

66. What are the appropriate null and alternative hypotheses?
   A. \( H_0: \mu_1 - \mu_2 = 0 \) and \( H_a: \mu_1 - \mu_2 \neq 0 \)
   B. \( H_0: \mu_1 - \mu_2 \neq 0 \) and \( H_a: \mu_1 - \mu_2 = 0 \)
   C. \( H_0: \mu_1 - \mu_2 = 0 \) and \( H_a: \mu_1 - \mu_2 < 0 \)
   D. \( H_0: \mu_1 - \mu_2 = 0 \) and \( H_a: \mu_1 - \mu_2 > 0 \)

   KEY: C

67. What is the value of the test statistic?
   A. \( t = -1.80 \)
   B. \( t = -2.00 \)
   C. \( t = -2.33 \)
   D. None of the above

   KEY: C

68. What is the p-value?
   A. \( p\text{-value} = 0.021 \)
   B. \( p\text{-value} = 0.022 \)
   C. \( p\text{-value} = 0.038 \)
   D. None of the above

   KEY: B
69. For a significance level of $\alpha = 0.05$, are the results statistically significant?
   A. No, results are not statistically significant because the $p$-value $< 0.05$.
   B. Yes, results are statistically significant because the $p$-value $< 0.05$.
   C. No, results are not statistically significant because the $p$-value $> 0.05$.
   D. Yes, results are statistically significant because the $p$-value $> 0.05$.

   KEY: B

70. Which of the following is an appropriate conclusion?
   A. The average delay for Airline 1 does appear to be less than the average delay for Airline 2.
   B. The results are not statistically significant so there is not enough evidence to conclude the average delay for Airline 1 is less than the average delay for Airline 2.
   C. The average delay for Airline 1 is at least 3 minutes less than Airline 2.
   D. None of the above.

   KEY: A

Questions 71 to 75: A bank official wants to assess if the average delays for two airlines are different. Let $\mu_1 =$ average delay for Airline 1 and $\mu_2 =$ average delay for Airline 2. A random sample of 10 flights for Airline 1 shows an average of 6 minutes delay with a standard deviation of 5 minutes. A random sample of 10 flights for Airline 2 shows an average of 12 minutes delay with a standard deviation of 5 minutes. Assume delay times are normally distributed, but do not assume the population variances are equal. Use the conservative “by hand” estimate for the degrees of freedom.

71. What are the appropriate null and alternative hypotheses?
   A. $H_0: \mu_1 - \mu_2 = 0$ and $H_a: \mu_1 - \mu_2 \neq 0$
   B. $H_0: \mu_1 - \mu_2 \neq 0$ and $H_a: \mu_1 - \mu_2 = 0$
   C. $H_0: \mu_1 - \mu_2 = 0$ and $H_a: \mu_1 - \mu_2 < 0$
   D. $H_0: \mu_1 - \mu_2 = 0$ and $H_a: \mu_1 - \mu_2 > 0$

   KEY: A

72. What is the value of the test statistic?
   A. $t = -2.68$
   B. $t = -1.50$
   C. $t = 1.50$
   D. $t = 2.68$

   KEY: A

73. For a significance level of 0.05, what is the critical value?
   A. The critical value = 1.83.
   B. The critical value = 2.10.
   C. The critical value = 2.26.
   D. None of the above.

   KEY: C

74. Are the results statistically significant?
   A. No, results are not statistically significant because $|t| < 1.83$.
   B. Yes, results are statistically significant because $|t| > 2.26$.
   C. No, results are not statistically significant because $|t| > 2.26$.
   D. None of the above.

   KEY: B

274
Questions 76 to 80: A researcher wants to assess if the average age when women first marry has increased from 1960 to 1990. Let \( \mu_1 = \) average age of first marriage for women in 1990 and \( \mu_2 = \) average age of first marriage for women in 1960. A random sample of 10 women married in 1990 showed an average age at marriage of 24.95 years, with a sample standard deviation of 2 years. A random sample of 20 women married in 1960 showed an average age at marriage of 23.1 years, with a sample standard deviation of 1.5 years. Assume that age of first marriage for women is normally distributed, but do not assume the population variances are equal. Use the conservative “by hand” estimate for the degrees of freedom.

76. What are the appropriate null and alternative hypotheses?
   A. \( H_0: \mu_1 - \mu_2 = 0 \) and \( H_a: \mu_1 - \mu_2 \neq 0 \)
   B. \( H_0: \mu_1 - \mu_2 \neq 0 \) and \( H_a: \mu_1 - \mu_2 = 0 \)
   C. \( H_0: \mu_1 - \mu_2 = 0 \) and \( H_a: \mu_1 - \mu_2 < 0 \)
   D. \( H_0: \mu_1 - \mu_2 = 0 \) and \( H_a: \mu_1 - \mu_2 > 0 \)
   KEY: D

77. What is the value of the test statistic?
   A. \( t = 1.80 \)
   B. \( t = 2.00 \)
   C. \( t = 2.33 \)
   D. \( t = 2.58 \)
   KEY: D

78. What is the \( p \)-value?
   A. \( p \)-value = 0.015
   B. \( p \)-value = 0.022
   C. \( p \)-value = 0.038
   D. None of the above
   KEY: A

79. For a significance level of \( \alpha = 0.05 \), are the results statistically significant?
   A. No, results are not statistically significant because the \( p \)-value < 0.05.
   B. Yes, results are statistically significant because the \( p \)-value < 0.05.
   C. No, results are not statistically significant because the \( p \)-value > 0.05.
   D. Yes, results are statistically significant because the \( p \)-value > 0.05.
   KEY: B

80. Which of the following is an appropriate conclusion?
   A. The average age of first marriage for women in 1990 does appear to be greater than the average age in 1960.
   B. The results are not statistically significant so there is not enough evidence to conclude that the average age in 1990 is greater than the average age in 1960.
   C. The average age of marriage is at least 23 years old for both 1960 and 1990.
   D. None of the above.
   KEY: A
Questions 81 to 85: A researcher wants to assess if there is a difference in the average age of onset of a certain disease for men and women who get the disease. Let $\mu_1 =$ average age of onset for women and $\mu_2 =$ average age of onset for men. A random sample of 30 women with the disease showed an average age of onset of 83 years, with a sample standard deviation of 11.5 years. A random sample of 20 men with the disease showed an average age of onset of 77 years, with a sample standard deviation of 4.5 years. Assume that ages at onset of this disease are normally distributed for each gender, do not assume the population variances are equal, and use the conservative “by hand” estimate for the degrees of freedom.

81. What are the appropriate null and alternative hypotheses?
   A. $H_0: \mu_1 - \mu_2 = 0$ and $H_a: \mu_1 - \mu_2 \neq 0$
   B. $H_0: \mu_1 - \mu_2 \neq 0$ and $H_a: \mu_1 - \mu_2 = 0$
   C. $H_0: \mu_1 - \mu_2 = 0$ and $H_a: \mu_1 - \mu_2 < 0$
   D. $H_0: \mu_1 - \mu_2 = 0$ and $H_a: \mu_1 - \mu_2 > 0$

   KEY: A

82. What is the value of the test statistic?
   A. $t = 2.00$
   B. $t = 2.33$
   C. $t = 2.58$
   D. None of the above

   KEY: C

83. What is the $p$-value?
   A. $p$-value = 0.009
   B. $p$-value = 0.018
   C. $p$-value = 0.015
   D. None of the above

   KEY: B

84. For a significance level of $\alpha = 0.05$, are the results statistically significant?
   A. No, results are not statistically significant because the $p$-value < 0.05.
   B. Yes, results are statistically significant because the $p$-value < 0.05.
   C. No, results are not statistically significant because the $p$-value > 0.05
   D. Yes, results are statistically significant because the $p$-value > 0.05.

   KEY: B

85. Which of the following is an appropriate conclusion?
   A. There is a statistically significant difference in average age at onset for men and women.
   B. The difference is not statistically significant: there is not enough evidence to conclude the average ages at onset are different.
   C. The average age at onset for women is 6 years later than the average age at onset for men.
   D. None of the above.

   KEY: A
Questions 86 to 90: A researcher wants to assess if there is a difference in the average life spans between men and women in Japan. Let $\mu_1 =$ average life span of Japanese women and $\mu_2 =$ average life span of Japanese men. A random sample of 10 women showed an average lifespan of 83 years, with a sample standard deviation of 7 years. A random sample of 10 men showed an average lifespan of 77 years, with a sample standard deviation of 6.4 years. Assume that life spans are normally distributed and that the population variances are equal.

86. What are the appropriate null and alternative hypotheses?
   A. $H_0$: $\mu_1 - \mu_2 = 0$ and $H_a$: $\mu_1 - \mu_2 \neq 0$
   B. $H_0$: $\mu_1 - \mu_2 \neq 0$ and $H_a$: $\mu_1 - \mu_2 = 0$
   C. $H_0$: $\mu_1 - \mu_2 = 0$ and $H_a$: $\mu_1 - \mu_2 < 0$
   D. $H_0$: $\mu_1 - \mu_2 = 0$ and $H_a$: $\mu_1 - \mu_2 > 0$

   KEY: A

87. What is the value of the test statistic?
   A. $t = 1.65$
   B. $t = 1.80$
   C. $t = 2.00$
   D. None of the above

   KEY: C

88. What is the $p$-value?
   A. $p$-value = 0.030
   B. $p$-value = 0.032
   C. $p$-value = 0.060
   D. None of the above

   KEY: C

89. For a significance level of $\alpha = 0.05$, are the results statistically significant?
   A. No, results are not statistically significant because the $p$-value < 0.05.
   B. Yes, results are statistically significant because the $p$-value < 0.05.
   C. No, results are not statistically significant because the $p$-value > 0.05
   D. Yes, results are statistically significant because the $p$-value > 0.05.

   KEY: C

90. Which of the following is an appropriate conclusion?
   A. There is a statistically significant difference in average life span between men and women.
   B. The difference is not statistically significant so there is not enough evidence to conclude that the life spans are different.
   C. The average life span of women is 6 years longer than the average life span of men.
   D. None of the above.

   KEY: B
Questions 91 to 96: Conscientiousness is a tendency to show self-discipline, act dutifully, and aim for achievement. The trait shows a preference for planned rather than spontaneous behavior. A random sample of 650 students is asked to fill out the Hogan Personality Inventory (HPI) to measure their level of conscientiousness. The 300 undergraduate students scored an average of 148 with a standard deviation of 16. The 350 graduate students had a mean score of 153 with a standard deviation of 21. Is this enough evidence to conclude that graduate students score higher, on average, on the HPI than undergraduate students?

91. What are the appropriate null and alternative hypotheses?
   A. $H_0: \mu_G - \mu_{UG} = 0$ and $H_a: \mu_G - \mu_{UG} \neq 0$
   B. $H_0: \mu_G - \mu_{UG} = 0$ and $H_a: \mu_G - \mu_{UG} > 0$
   C. $H_0: \mu_G - \mu_{UG} = 0$ and $H_a: \mu_G - \mu_{UG} < 0$
   D. $H_0: \mu_G - \mu_{UG} \neq 0$ and $H_a: \mu_G - \mu_{UG} = 0$

   KEY: B

92. What is the value of the pooled sample variance for the difference between the two sample means?
   A. 355.64
   B. 18.86
   C. 1.48
   D. 16.15

   KEY: A

93. What is the value of the test statistic for testing these hypotheses using a pooled two sample $t$-test?
   A. 0.27
   B. 3.37
   C. 3.44
   D. 6.89

   KEY: B

94. Are the results significant using a significance level of 5%?
   A. Yes, because the $p$-value < 0.05.
   B. Yes, because the $p$-value > 0.05.
   C. No, because the $p$-value < 0.05.
   D. No, because the $p$-value > 0.05.

   KEY: A

95. Which of the following is an appropriate conclusion?
   A. The population mean HPI score of graduate students is 5 points higher than the population mean HPI score of undergraduate students.
   B. The mean HPI score of graduate students is not significantly higher than the mean HPI score of undergraduate students.
   C. The mean HPI score of graduate students is significantly higher than the mean HPI score of undergraduate students.
   D. The mean HPI score of graduate students is higher than the mean HPI score of undergraduate students.

   KEY: C
Questions 96 to 99: Spatial perception is measured on a scale from 0 to 10. Two groups of 8th grade children are tested for spatial perception. The students in group 1 were first given a short course on spatial concepts. The students in group 2 received no such instruction. SPSS was used to calculate descriptive statistics for the two samples.

| Group Statistics |
|------------------|---------|----------|-----------|
|                  | Group 1 | Group 2  |
| Spatial Perception| 8       | 10       |
| N                | 7.13    | 5.10     |
| Mean             | 1.126   | 1.792    |
| Std. Deviation   | .398    | .567     |
| Std. Error Mean  |         |          |

96. What is the value of the pooled standard error?
   A. 0.483
   B. 0.693
   C. 1.537
   D. 0.729

KEY: D

97. What is the value of the test statistic for the pooled two sample t-test?
   A. 1.39
   B. 2.79
   C. 2.93
   D. 4.21

KEY: B

98. What do we know about the p-value if we wish to test if participation in the course results in significantly higher spatial perception scores?
   A. p-value < 0.01
   B. 0.01 < p-value < 0.025
   C. 0.025 < p-value < 0.05
   D. p-value > 0.05

KEY: A

99. Which of the following is an appropriate conclusion?
   A. The population mean spatial perception score of children who take the course on spatial concepts is more than 2 points higher than the population mean score of children who don’t take the course.
   B. The mean spatial perception score of children who take the course on spatial concepts is significantly higher than the average score of children who don’t take the course.
   C. The mean spatial perception score of children who take the course on spatial concepts is not significantly higher than the average score of children who don’t take the course.
   D. The mean spatial perception score of children who take the course on spatial concepts is higher than the average score of children who don’t take the course.

KEY: B
Questions 100 and 101: For the variable “Time spent watching TV in Typical Day,” here are results of a two-sample $t$-procedure that compares a random sample of men and women at a college.

<table>
<thead>
<tr>
<th>Sex</th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>SE Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>f</td>
<td>116</td>
<td>1.95</td>
<td>1.51</td>
<td>0.14</td>
</tr>
<tr>
<td>m</td>
<td>59</td>
<td>2.37</td>
<td>1.87</td>
<td>0.24</td>
</tr>
</tbody>
</table>

$95\%$ CI for $\mu_f - \mu_m$: ($-0.97$, $0.14$)

$T$-Test: $\mu_f = \mu_m$ ($vs$ not $=$): $T = -1.49$ $P = 0.14$ $DF = 97$

100. What is the alternative hypothesis of the $t$-test?
A. $H_a$: $\mu_1 - \mu_2 = 0$
B. $H_a$: $\mu_1 - \mu_2 \neq 0$
C. $H_a$: $\mu_1 - \mu_2 > 0$
D. $H_a$: $\mu_1 - \mu_2 < 0$

KEY: B

101. Which of the following is the correct conclusion about these results using a 5% significance level?
A. The mean TV watching times of men and women at the college are equal.
B. There is a statistically significant difference between the mean TV watching times of men and women at the college.
C. There is not a statistically significant difference between the mean TV watching times of men and women at the college.
D. There is not enough information to judge statistical significance here.

KEY: C

Questions 102 to 104: In the survey of a random sample of students at a university, two questions were “How many hours per week do you usually study?” and “Have you smoked marijuana in the past six months?” An analysis of the results produced the following Minitab output:

<table>
<thead>
<tr>
<th>Marijan</th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>SE Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>49</td>
<td>12.41</td>
<td>6.95</td>
<td>0.99</td>
</tr>
<tr>
<td>No</td>
<td>150</td>
<td>15.53</td>
<td>8.79</td>
<td>0.72</td>
</tr>
</tbody>
</table>

Difference = $\mu_1 (Yes) - \mu_2 (No)$
Estimate for difference: $-3.12$
$95\%$ CI for difference: ($-5.55$, $-0.69$)

A research question of interest is whether students who have smoked marijuana (group 1) in the past 6 months study fewer hours on average per week than those who have not (group 2).

102. What is the appropriate alternative hypothesis for this question?
A. $H_a$: $\mu_1 - \mu_2 = 0$
B. $H_a$: $\mu_1 - \mu_2 < 0$
C. $H_a$: $\mu_1 - \mu_2 \neq 0$
D. $H_a$: $\mu_1 - \mu_2 < 0$

KEY: D
103. Which of the following is the appropriate unpooled test statistic for determining if there is a difference in average study hours per week for the two groups?

A. \[ t = \frac{12.41 - 15.53}{\sqrt{0.99 + 0.72}} \]

B. \[ t = \frac{12.41 - 15.53}{\sqrt{6.95 + 8.79}} \]

C. \[ t = \frac{12.41 - 15.53}{\sqrt{\frac{6.95}{49} + \frac{8.79}{150}}} \]

D. \[ t = \frac{12.41 - 15.53}{\sqrt{\frac{(6.95)^2}{49} + \frac{(8.79)^2}{150}}} \]

KEY: D

104. Based on the information given in the output, what conclusion can be made about the difference in time spent studying for the two groups?

A. There is a statistically significant difference.

B. There is not a statistically significant difference.

C. It is impossible to know if there is a statistically significant difference because no p-value is provided.

D. It is impossible to know if there is a statistically significant difference because the test is one-sided and the information provided is two-sided.

KEY: A
Questions 105 to 109: A large car insurance company is conducting a study to see if male and female drivers have the same number of accidents, on average, or if male drivers (who tend to be thought of as more aggressive drivers) have more. Data on the number of accidents in the past 5 years is collected for randomly selected drivers who are insured by this company. An analysis of the results produced the following output.

<table>
<thead>
<tr>
<th>Group Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex of insured</td>
</tr>
<tr>
<td>Number of accidents past 5 years</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Independent Samples Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Levene's Test for Equality of Variances</td>
</tr>
<tr>
<td>F</td>
</tr>
<tr>
<td>Number of accidents past 5 years</td>
</tr>
<tr>
<td>Equal variances not assumed</td>
</tr>
</tbody>
</table>

105. State the appropriate null and alternative hypotheses and be sure to identify the corresponding parameters.

KEY: \( H_0: \mu_1 - \mu_2 = 0 \) and \( H_a: \mu_1 - \mu_2 > 0 \) where \( \mu_1 \) is the population mean number of accidents for male drivers and \( \mu_2 \) is the population mean number of accidents for female drivers.

106. Since the sample standard deviations are fairly similar, the pooled \( t \)-test will be used (and thus the results can be found in the Equal variances assumed row of output). What is the value of the pooled \( t \)-test statistic?

KEY: The pooled \( t \)-test statistic is \( t = 2.405 \).

107. Note that the \( p \)-value for the two-sided test is \( p = 0.0193 \). What would be the \( p \)-value for the one-sided test of interest here?

KEY: Since the test statistic is positive and the alternative hypothesis was ‘\( > \)’, the \( p \)-value would be \( 0.0193/2 = 0.00965 \).

108. Are the results statistically significant at the 5% significance level?

KEY: Yes, since the \( p \)-value is much less than 0.05, we would reject the null hypothesis and state that the results are statistically significant.

109. What would be the appropriate conclusion using a 5% significance level?

KEY: There is sufficient evidence to conclude that the average number of accidents for male drivers is higher than that for female drivers.
Sections 13.5 – 13.6

110. In each of the following cases, we wish to test the null hypothesis \( H_0: \mu = 10 \) vs. \( H_a: \mu \neq 10 \). In which case can this null hypothesis not be rejected at a significance level \( \alpha = 0.05 \)?
   A. 90% confidence interval for \( \mu \) is (9 to 12).
   B. 95% confidence interval for \( \mu \) is (11 to 21).
   C. 98% confidence interval for \( \mu \) is (4.4 to 9.4).
   D. 99% confidence interval for \( \mu \) is (5 to 9).
   KEY: A

111. In each of the following cases, we wish to test the null hypothesis \( H_0: \mu = 18 \). In which case can this null hypothesis be rejected at a significance level \( \alpha = 0.05 \)?
   A. 95% confidence interval for \( \mu \) is (16 to 21), \( H_a: \mu \neq 18 \).
   B. 90% confidence interval for \( \mu \) is (15 to 20), \( H_a: \mu < 18 \).
   C. 90% confidence interval for \( \mu \) is (19 to 26), \( H_a: \mu > 18 \).
   D. 90% confidence interval for \( \mu \) is (15 to 22), \( H_a: \mu \neq 18 \).
   KEY: C

112. Each of the following presents a two-sided confidence interval and the alternative hypothesis of a corresponding hypothesis test. In which case can we not use the given confidence interval to make a decision at a significance level \( \alpha = 0.05 \)?
   A. 95% confidence interval for \( p_1 - p_2 \) is (−0.15 to 0.07), \( H_a: p_1 - p_2 \neq 0 \).
   B. 95% confidence interval for \( p \) is (.12 to .28), \( H_a: p > 0.10 \).
   C. 90% confidence interval for \( \mu \) is (101 to 105), \( H_a: \mu \neq 100 \).
   D. 90% confidence interval for \( \mu_1 - \mu_2 \) is (3 to 15), \( H_a: \mu_1 - \mu_2 > 0 \).
   KEY: C

113. A random sample of 25 college males was obtained and each was asked to report their actual height and what they wished as their ideal height. A 95% confidence interval for \( \mu_d = \text{average difference between their ideal and actual heights} \) was 0.8” to 2.2”. Based on this interval, which one of the null hypotheses below (versus a two-sided alternative) can be rejected?
   A. \( H_0: \mu_d = 0.5 \)
   B. \( H_0: \mu_d = 1.0 \)
   C. \( H_0: \mu_d = 1.5 \)
   D. \( H_0: \mu_d = 2.0 \)
   KEY: A

114. As Internet usage flourishes, so do questions of security and confidentiality of personal information. A survey of U.S. adults resulted in a 95% confidence interval for the proportion of all U.S. adults who would never give personal information to a company of (0.22, 0.30). Based on this interval, which one of the null hypotheses below (versus a two-sided alternative) can be rejected?
   A. \( H_0: p = 2/7 \)
   B. \( H_0: p = 1/3 \)
   C. \( H_0: p = 1/4 \)
   D. \( H_0: p = 0.26 \)
   KEY: B
Questions 115 to 120: A small bakery is trying to predict how many loaves of bread to bake daily. They randomly sample daily sales records from the past year. In these days, the bakery sold between 40 and 80 loaves per day with a 95% confidence interval for the population mean daily loaf demand given by (51, 54).

115. What was the mean number of loaves sold daily for the sample of 50 days?
   A. 50
   B. 60
   C. 52.5
   D. Cannot be determined.
   KEY: C

116. How would a 99% confidence interval compare to the 95% confidence interval?
   A. It would be wider.
   B. It would be narrower.
   C. It would be the same width.
   D. Cannot be determined.
   KEY: A

117. Based on this interval, what is the decision for testing $H_0: \mu = 50$ versus $H_a: \mu \neq 50$ at the 5% significance level?
   A. Reject the null hypothesis.
   B. Fail to reject the null hypothesis.
   C. Cannot be determined.
   KEY: A

118. Based on this interval, what is the decision for testing $H_0: \mu = 50$ versus $H_a: \mu \neq 50$ at the 10% significance level?
   A. Reject the null hypothesis.
   B. Fail to reject the null hypothesis.
   C. Cannot be determined.
   KEY: A

119. Based on this interval, what is the decision for testing $H_0: \mu = 52$ versus $H_a: \mu \neq 52$ at the 5% significance level?
   A. Reject the null hypothesis.
   B. Fail to reject the null hypothesis.
   C. Cannot be determined.
   KEY: B

120. Based on this interval, what is the decision for testing $H_0: \mu = 52$ versus $H_a: \mu \neq 52$ at the 10% significance level?
   A. Reject the null hypothesis.
   B. Fail to reject the null hypothesis.
   C. Cannot be determined.
   KEY: C
121. The average time in years to get an undergraduate degree in computer science was compared for men and women. Random samples of 100 male computer science majors and 100 female computer science majors were taken. Choose the appropriate parameter(s) for this situation.
   A. \( p \)
   B. \( p_1 - p_2 \)
   C. \( \mu \)
   D. \( \mu_1 - \mu_2 \)
   KEY: D

122. At entrance C of the football stadium there are 3 swing gates for the spectators. On football Saturday, two high school students sit at this entrance and count the number of people who enter through each of the swing gates. They also write down the gender of these spectators. On Monday, during their statistics class, they wish to determine if men are more likely to enter through the middle swing gate than women. What is the appropriate parameter in this situation?
   A. \( \mu_d \)
   B. \( p_1 - p_2 \)
   C. \( p \)
   D. \( \mu_1 - \mu_2 \)
   KEY: C

123. According to the CIRP Freshman Survey, UCLA’s annual survey of the nation’s entering students at four-year colleges and universities, conducted in 2010, female students were far less likely to report high levels of emotional health than male students. The difference was reported as statistically significant. Choose the appropriate parameter(s) for this situation.
   A. \( \mu \)
   B. \( \mu_d \)
   C. \( \mu_1 - \mu_2 \)
   D. \( p_1 - p_2 \)
   KEY: D

124. For what problem is a one-sample \( t \)-test used?
   A. To test a hypothesis about a proportion.
   B. To test a hypothesis about a mean.
   C. To test a hypothesis about the difference between two means for independent samples.
   D. To test a hypothesis about the difference between two proportions for independent samples.
   KEY: B

125. For what problem is a two-sample \( t \)-test used?
   A. To test a hypothesis about a mean.
   B. To test a hypothesis about the mean difference for paired data.
   C. To test a hypothesis about the difference between two means for independent samples.
   D. To test a hypothesis about the difference between two proportions for independent samples.
   KEY: C
126. A sociologist has been studying college instructors and is especially interested in those instructors who teach night classes. Amongst others, he collected data on the course grade from the student evaluations and the highest level of education the instructor has completed (Master or Ph.D.). If the sociologist wishes to determine how much higher the course grades are for courses taught by an instructor with a Ph.D. than those taught by an instructor with a master’s degree, which inference procedure should he use?

A. A t-test about the difference between two means for independent samples.
B. A t-test about the mean difference for paired data.
C. A confidence interval for the difference between two means for independent samples.
D. A confidence interval for the mean difference for paired data.

KEY: C

127. One hundred adults who had either Type 1 or Type 2 diabetes were divided into 2 groups to determine if the inhalable version of insulin is able to manage blood sugar levels just as well as injected insulin. The first group used the inhaler for 3 months first and then the injections for 3 months. The other group used the two methods in reverse order. If we wish to determine if there is a difference in average blood sugar levels using the two different methods, which inference procedure should we use?

A. A t-test about the difference between two means for independent samples.
B. A t-test about the mean difference for paired data.
C. A confidence interval for the difference between two means for independent samples.
D. A confidence interval for the mean difference for paired data.

KEY: B

128. One hundred adults who had either Type 1 or Type 2 diabetes were divided into 2 groups to determine if the inhalable version of insulin is able to manage blood sugar levels just as well as injected insulin. The first group received the inhaler; the second group continued using their injections. If we wish to determine if there is a difference in average blood sugar levels at the end of the trial period between the inhaler group and the injection group, which inference procedure should we use?

A. A t-test about the difference between two means for independent samples.
B. A t-test about the mean difference for paired data.
C. A confidence interval for the difference between two means for independent samples.
D. A confidence interval for the mean difference for paired data.

KEY: A

129. One hundred adults who had either Type 1 or Type 2 diabetes were divided into 2 groups to determine if the inhalable version of insulin is able to manage blood sugar levels just as well as injected insulin. At the end of the trial period we count the number of Type 1 diabetics who reacted better to the inhaler than to the injections. We do the same for the Type 2 diabetics. If we wish to determine if there is a difference in the percentages of diabetics who reacted better to the inhaler between Type 1 and Type 2 diabetics, which inference procedure should we use?

A. A t-test about the difference between two means for independent samples.
B. A z-test about the difference between two proportions.
C. A confidence interval for the difference between two means for independent samples.
D. A confidence interval for the difference between two proportions.

KEY: B
130. A biologist wants to measure the effect of a fertilizer on the growth of roses. She plants 20 roses in 10 plots of land with 2 roses in each plot. She randomly applies the fertilizer to one of the two roses in each plot, and then compares the average height of the roses with fertilizer to the average height of roses without fertilizer after 1 month. What is the appropriate statistical procedure she should use?

KEY: The study design is a paired design, so a paired \( t \)-procedure could be used, provided the assumption of normality for the height of roses is reasonable. She would probably want to compute a confidence interval for the average difference in the population of roses like these. She might also want to test to see if the population mean difference is greater than 0, with fertilized roses growing taller.

131. A researcher wants to assess if the time of day during which one works affects job satisfaction. A study compared job satisfaction for day-shift nurses versus night-shift nurses. Independent samples of nurses from each group were obtained and their job satisfaction was measured on a numerical scale with higher scores indicating greater satisfaction. Identify the appropriate parameter for this situation and name an appropriate technique that might be performed provided the underlying assumptions are met.

KEY: Since the response is quantitative and there are two independent samples, the appropriate parameter would be the difference between two population means \( \mu_1 - \mu_2 \) where \( \mu_1 \) is the mean job satisfaction for day-shift nurses and \( \mu_2 \) is the mean job satisfaction for night-shift nurses. A potential test is the two independent samples \( t \)-test.

132. A study will be conducted to compare the proportion of earth-quake insured residents between two counties. Surveys will be mailed to a sample of residents in each county. The results will be used to assess if the county that is closer to a major earthquake fault has a higher proportion of residents with such insurance. Identify the appropriate parameter for this situation and name an appropriate technique that might be performed provided the underlying assumptions are met.

KEY: Since the response is categorical and there are two independent samples, the appropriate parameter would be the difference between two population proportions \( p_1 - p_2 \) where \( p_1 \) is the proportion of residents with such insurance in county 1 and \( p_2 \) is the proportion of residents with such insurance in county 2. A two independent samples \( z \)-test would be performed provided the samples are representative and the sample sizes are sufficiently large.

133. A company is interested in estimating the mean number of days of sick leave taken by all its employees. The firm’s statistician selects a random sample of personnel files, notes the number of sick days taken by each employee over previous year. Identify the appropriate parameter for this situation and name an appropriate technique that might be performed provided the underlying assumptions are met.

KEY: Since the response is quantitative and there is one sample, appropriate parameter would be the population mean \( \mu \), the mean number of days of sick leave for the population of all employees. A potential technique is a one-sample \( t \) confidence interval for estimating the population mean.
**Questions 134 to 139:** Unoccupied seats on flights cause airlines to lose revenue. A large airline wanted to estimate its average number of unoccupied seats per flight. Based on a random sample of flights, the 95% confidence interval for mean number of unoccupied seats per flight was given as: (7.5, 9.5).

134. What was the mean number of unoccupied seats per flight for the sample of flights?
   **KEY:** The sample mean would be the midpoint of the confidence interval, so $\bar{x} = 8.5$ seats.

135. Based on the 95% confidence interval, what would be the decision for testing $H_0: \mu = 7$ versus $H_a: \mu \neq 7$ at the 5% significance level?
   **KEY:** Since the 95% confidence interval does not contain the value of 7 seats, we would reject the null hypothesis.

136. What would be the decision for testing $H_0: \mu = 7$ versus $H_a: \mu \neq 7$ at the 10% significance level?
   **KEY:** Since the 90% confidence interval would be narrower than the 95% interval and still would not contain the value of 7 seats, we would reject the null hypothesis.

137. Based on the 95% confidence interval, what would be the decision for testing $H_0: \mu = 9$ versus $H_a: \mu \neq 9$ at the 5% significance level?
   **KEY:** Since the 95% confidence interval does contain the value of 9 seats, we would fail to reject the null hypothesis.

138. What would be the decision for testing $H_0: \mu = 9$ versus $H_a: \mu \neq 9$ at the 1% significance level?
   **KEY:** Since the 99% confidence interval would be wider than the 95% interval and thus still would contain the value of 9 seats, we would fail to reject the null hypothesis.

139. What would be the decision for testing $H_0: \mu = 9$ versus $H_a: \mu \neq 9$ at the 10% significance level?
   **KEY:** The 90% confidence interval would be narrower than the 95% interval and without more information, we cannot determine whether or not the 90% interval would contain the value of 9 seats, so a decision cannot be made without more information.
Sections 13.7 – 13.8

140. In a statistical test, define the events: A = the null hypothesis is true and B = the null hypothesis is rejected. Using that notation and notation on conditional probability from Chapter 7, the power of a test can be written as:

A. \( P(A|B) \)
B. \( P(B|A) \)
C. \( P(B|\text{not } A) \)
D. \( P(A|\text{not } B) \)

KEY: C

141. An agricultural experiment will be done to compare the effectiveness of two different soil treatments. Which of the following sample sizes would lead to a test with the lowest power?

A. \( n = 400 \)
B. \( n = 100 \)
C. \( n = 20 \)
D. \( n = 10 \)

KEY: D

142. An experiment will be conducted to compare the effectiveness of a new treatment against a placebo. Which of the following combinations would lead to a test with the highest power?

A. \( n_1 = 40, n_2 = 45 \) and \( \mu_1 - \mu_2 = 10 \)
B. \( n_1 = 40, n_2 = 45 \) and \( \mu_1 - \mu_2 = 20 \)
C. \( n_1 = 70, n_2 = 80 \) and \( \mu_1 - \mu_2 = 10 \)
D. \( n_1 = 70, n_2 = 80 \) and \( \mu_1 - \mu_2 = 20 \)

KEY: D

143. A study compared the average number of courses taken by a random sample of freshmen at a university with the average number of courses taken by a separate random sample of freshmen at a community college. Which of the following combinations would lead to a test with the highest power?

A. \( n_1 = 50, n_2 = 100 \) and \( \mu_1 - \mu_2 = 1.2 \)
B. \( n_1 = 70, n_2 = 120 \) and \( \mu_1 - \mu_2 = 1.2 \)
C. \( n_1 = 50, n_2 = 100 \) and \( \mu_1 - \mu_2 = 0.8 \)
D. \( n_1 = 70, n_2 = 120 \) and \( \mu_1 - \mu_2 = 0.8 \)

KEY: B

144. When the alternative hypothesis is true, which one of these statements about the power of a significance test is correct?

A. The smaller the difference between the true and null values of a parameter, the larger the power.
B. The power of a test is the probability of a Type 1 error
C. The larger the sample size, the smaller the power of significance test.
D. The power of a test is the probability of not making a Type 2 error

KEY: D

145. If the word significant is used to describe a result in a news article reporting on a study,

A. the \( p \)-value for the test must have been very large.
B. the effect size must have been very large.
C. the sample size must have been very small.
D. it may be used in the everyday sense and not in the statistical sense.

KEY: D
**Questions 146 and 147:** Cholesterol measurements for a sample of males following a new diet plan gave a mean of 208 and a standard deviation of 10.

146. What is the estimated effect size for comparing this result to the 200 level?
   - A. 1.8
   - B. 8
   - C. 0.8
   - D. -0.8

KEY: C

147. Based on Cohen’s interpretation, this effect is
   - A. a small effect, not obvious without statistics.
   - B. a medium effect, obvious to careful observers.
   - C. a large effect, obvious to most observers.

KEY: C

**Questions 148 and 149:** A sample of passengers flying on small planes gave a mean total weight (passenger weight including shoes, clothes, and carry-on) of 200 pounds and a standard deviation of 60 pounds.

148. What is the estimated effect size for comparing this result to the FDA guideline level of 185 pounds?
   - A. 0.25
   - B. -0.25
   - C. -15
   - D. 15

KEY: A

149. Based on Cohen’s interpretation, this effect is
   - A. a small effect, not obvious without statistics.
   - B. a medium effect, obvious to careful observers.
   - C. a large effect, obvious to most observers.

KEY: A

**Questions 150 and 151:** A pooled two independent samples t-test was conducted to compare the average number of accidents for male drivers to that for female drivers. The t-test statistic was 2.4 and the sample sizes were 32 males and 30 females.

150. What is the estimated effect size for this study?
   - A. 2.4
   - B. 62
   - C. 0.04
   - D. 0.61

KEY: D

151. Based on Cohen’s interpretation, this effect is
   - A. a small effect, not obvious without statistics.
   - B. a medium effect, obvious to careful observers.
   - C. a large effect, obvious to most observers.

KEY: B
152. What is the estimated effect size for a test for the difference between two population means with independent samples with sizes of 25 and 35 respectively and a pooled t-test statistic of $t = -2.34$?

A. 0.23  
B. -0.43  
C. -0.61  
D. -0.16  

KEY: C

153. Compute the effect size in a one-sample t-test with a sample size of $n = 66$ and a test statistic of $t = 2.27$. How is the effect size interpreted based on Cohen’s guidelines?

A. 0.28, a small effect  
B. 0.38, a medium effect  
C. 0.03, a small effect  
D. 0.63, a large effect  

KEY: A

154. A random sample of 25 third graders scored an average of 3.4 on a standardized reading test. The standard deviation was 0.95. A test is conducted to determine if the mean score is significantly higher than 3. What is the effect size of this test?

A. 0.42  
B. 2.11  
C. 3.16  
D. None of the above  

KEY: A

**Questions 155 and 156:** A group of women suffering from various types of obsessive fear is given a therapeutic course to learn how to handle their fear. At the end of the course the women are given a questionnaire to determine how afraid they are. Six months after the course the women answer the questionnaire again as a follow-up measure. The fear-scores are typed into SPSS. Output of the paired samples t-test is shown below.

<table>
<thead>
<tr>
<th>Paired Differences</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pair 1 post - follow up</td>
<td>2.520</td>
<td>4.417</td>
<td>.883</td>
<td>2.853</td>
<td>24</td>
<td>0.0088</td>
</tr>
</tbody>
</table>

155. What is the effect size that should be reported with this test?

A. 0.88  
B. 0.57  
C. 0.0088  
D. None of the above  

KEY: B

156. Based on Cohen’s interpretation, this effect is

A. a small effect, not obvious without statistics.  
B. a medium effect, obvious to careful observers.  
C. a large effect, obvious to most observers.  

KEY: B
157. Explain what the effect size measures.

**KEY:** The effect size measures how much the truth differs from chance or from a controlled condition.

158. Explain the role of effect size in meta-analyses.

**KEY:** A meta-analysis is one in which the results on the same topic are statistically combined across various studies. Since the studies can differ on many levels including the measurement units and sample sizes, using effect sizes (which are not influenced by sample size or units) are effective.

159. Using Cohen’s interpretation of effect size, what would an effect size of $d = 0.12$ be interpreted as?

**KEY:** An effect size of $d = 0.12$ would be a small effect size, which would not be obvious without statistics.

160. Using Cohen’s interpretation of effect size, what would an effect size of $d = -0.52$ be interpreted as?

**KEY:** An effect size of $d = -0.52$ would be a medium effect size, which should be evident to a careful observer.

161. Using Cohen’s interpretation of effect size, what would an effect size of $d = 0.83$ be interpreted as?

**KEY:** An effect size of $d = 0.83$ would be a large effect size, which should be obvious to most observers.

162. Compute the effect size in a one-sample $t$-test with a sample size of $n = 25$ and a test statistic of $t = 2.00$.

**KEY:** The effect size is $d = 0.40$.

163. Compute the effect size in a paired $t$-test for the mean difference based on 100 paired observations and a test statistic of $t = -3.0$.

**KEY:** The effect size is $d = -0.30$.

164. Compute the effect size in a test for the difference between two population means with independent samples with sample of sizes of 50 each and a test statistic of $t = -2.0$.

**KEY:** The effect size is $d = -0.40$.

165. Suppose that the average age when women first marry was 23 years in 1960 and 25 years in 1990, with a standard deviation of 2 years in both 1960 and 1990. What is the effect size that measures the increase in age when women get married from 1960 to 1990?

**KEY:** The effect size = 1.

166. The headline of a newspaper reads: “More college students than expected use tobacco.” The article went on to say that a recent survey shows a significant increase in the use of tobacco among college students with about one-third of college students reporting having used tobacco in the past 30 days. Comment on what a person should consider in evaluating the quality of this reported study result.

**KEY:** There are a number of issues to consider. Is the use of the word significant being used in the everyday sense or in the statistical sense? The number of college students surveyed would also be of use in evaluating the quality of the results as well as how the survey was conducted.