8-3 Basic Skills and Concepts

Independent Samples and Matched Pairs. In Exercises 1–4, determine whether the samples are independent or consist of matched pairs.

1. The effectiveness of Prilosec for treating heartburn is tested by measuring gastric acid secretion in a group of patients treated with Prilosec and another group of patients given a placebo.

2. The effectiveness of Prilosec for treating heartburn is tested by measuring gastric acid secretion in patients before and after the drug treatment. The data consist of the before/after measurements for each patient.

3. The accuracy of verbal responses is tested in an experiment in which subjects report their weights and they are then weighed on a physician’s scale. The data consist of the reported weight and measured weight for each subject.

4. The effect of sugar as an ingredient is tested with a sample of cans of regular Coke and another sample of cans of diet Coke.

In Exercises 5–24, assume that the two samples are independent simple random samples selected from normally distributed populations. Do not assume that the population standard deviations are equal.

5. Hypothesis Test for Effect of Marijuana Use on College Students Many studies have been conducted to test the effects of marijuana use on mental abilities. In one such study, groups of light and heavy users of marijuana in college were tested for memory recall, with the results given below (based on data from “The Residual Cognitive Effects of Heavy Marijuana Use in College Students” by Pope and Yurgelun-Todd, Journal of the American Medical Association, Vol. 275, No. 7). Use a 0.01 significance level to test the claim that the population of heavy marijuana users has a lower mean than the light users. Should marijuana use be of concern to college students?

   Items sorted correctly by light marijuana users: \( n = 64, \bar{x} = 53.3, s = 3.6 \)
   Items sorted correctly by heavy marijuana users: \( n = 65, \bar{x} = 51.3, s = 4.5 \)

6. Confidence Interval for Effects of Marijuana Use on College Students Refer to the sample data used in Exercise 5 and construct a 98% confidence interval for the difference between the two population means. Does the confidence interval include zero? What does the confidence interval suggest about the equality of the two population means?

7. Confidence Interval for Bipolar Depression Treatment In clinical experiments involving different groups of independent samples, it is important that the groups be similar in the important ways that affect the experiment. In an experiment designed to test the effectiveness of paroxetine for treating bipolar depression, subjects were measured using the Hamilton depression scale with the results given below (based on data from “Double-Blind, Placebo-Controlled Comparison of Imipramine and Paroxetine...
in the Treatment of Bipolar Depression” by Nemeroff et al., *American Journal of Psychiatry*, Vol. 158, No. 6). Construct a 95% confidence interval for the difference between the two population means. Based on the results, does it appear that the two populations have different means? Should paroxetine be recommended as a treatment for bipolar depression?

Placebo group: \( n = 43, \bar{x} = 21.57, s = 3.87 \)
Paroxetine treatment group: \( n = 33, \bar{x} = 20.38, s = 3.91 \)

8. **Hypothesis Test for Bipolar Depression Treatment** Refer to the sample data in Exercise 7 and use a 0.05 significance level to test the claim that the treatment group and placebo group come from populations with the same mean. What does the result of the hypothesis test suggest about paroxetine as a treatment for bipolar depression?

9. **Hypothesis Test for Magnet Treatment of Pain** People spend huge sums of money (currently around $5 billion annually) for the purchase of magnets used to treat a wide variety of pains. Researchers conducted a study to determine whether magnets are effective in treating back pain. Pain was measured using the visual analog scale, and the results given below are among the results obtained in the study (based on data from “Bipolar Permanent Magnets for the Treatment of Chronic Lower Back Pain: A Pilot Study” by Collacott, Zimmerman, White, and Rindone, *Journal of the American Medical Association*, Vol. 283, No. 10). Use a 0.05 significance level to test the claim that those treated with magnets have a greater reduction in pain than those given a sham treatment (similar to a placebo). Does it appear that magnets are effective in treating back pain? Is it valid to argue that magnets might appear to be effective if the sample sizes are larger?

Reduction in pain level after magnet treatment: \( n = 20, \bar{x} = 0.49, s = 0.96 \)
Reduction in pain level after sham treatment: \( n = 20, \bar{x} = 0.44, s = 1.4 \)

10. **Confidence Interval for Magnet Treatment of Pain** Refer to the sample data from Exercise 9 and construct a 90% confidence interval estimate of the difference between the mean reduction in pain for those treated with magnets and the mean reduction in pain for those given a sham treatment. Based on the result, does it appear that the magnets are effective in reducing pain?

11. **Inferences from Samples of Regular Coke and Diet Coke** Using [Data Set 17 in Appendix B], we find the sample statistics for the weights (in pounds) of regular Coke and diet Coke as listed in the margin.

a. Use a 0.01 significance level to test the claim that cans of regular Coke and diet Coke have the same mean weight. If there appears to be a difference, try to provide an explanation.

b. Construct a 99% confidence interval estimate of \( \mu_1 - \mu_2 \), the difference between the mean weight of regular Coke and the mean weight of diet Coke.

12. **Cigarette Filters and Nicotine** Refer to the sample results listed in the margin for the measured nicotine contents of randomly selected filtered and nonfiltered king-size cigarettes. All measurements are in milligrams, and the data are from the Federal Trade Commission.

a. Use a 0.05 significance level to test the claim that king-size cigarettes with filters have a lower mean amount of nicotine than the mean amount of nicotine in nonfiltered king-size cigarettes.
b. Construct a 90% confidence interval estimate of the difference between the two population means.
c. Do cigarette filters appear to be effective in reducing nicotine?

13. **Confidence Interval for Identifying Psychiatric Disorders** Are severe psychiatric disorders related to biological factors that can be physically observed? One study used x-ray computed tomography (CT) to collect data on brain volumes for a group of patients with obsessive-compulsive disorders and a control group of healthy persons. Sample results for volumes (in mL) follow for the right cordate (based on data from “Neuroanatomical Abnormalities in Obsessive-Compulsive Disorder Detected with Quantitative X-Ray Computed Tomography,” by Luxenberg et al., *American Journal of Psychiatry*, Vol. 145, No. 9). Construct a 99% confidence interval estimate of the difference between the mean brain volume for the healthy control group and the mean brain volume for the obsessive-compulsive group. What does the confidence interval suggest about the difference between the two population means? Based on this result, does it seem that obsessive-compulsive disorders have a biological basis?

Control group: \( n = 10, \bar{x} = 0.45, s = 0.08 \)

Obsessive-compulsive patients: \( n = 10, \bar{x} = 0.34, s = 0.08 \)

14. **Hypothesis Test for Identifying Psychiatric Disorders** Refer to the sample data in Exercise 13 and use a 0.01 significance level to test the claim that there is a difference between the two population means. Based on the result, does it seem that obsessive-compulsive disorders have a biological basis?

15. **Confidence Interval for Effects of Alcohol** An experiment was conducted to test the effects of alcohol. The errors were recorded in a test of visual and motor skills for a treatment group of people who drank ethanol and another group given a placebo. The results are shown in the accompanying table (based on data from “Effects of Alcohol Intoxication on Risk Taking, Strategy, and Error Rate in Visuomotor Performance,” by Streufert et al., *Journal of Applied Psychology*, Vol. 77, No. 4). Construct a 95% confidence interval estimate of the difference between the two population means. Do the results support the common belief that drinking is hazardous for drivers, pilots, ship captains, and so on? Why or why not?

<table>
<thead>
<tr>
<th>Treatment Group</th>
<th>Placebo Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n_1 = 22 )</td>
<td>( n_2 = 22 )</td>
</tr>
<tr>
<td>( \bar{x}_1 = 4.20 )</td>
<td>( \bar{x}_2 = 1.71 )</td>
</tr>
<tr>
<td>( s_1 = 2.20 )</td>
<td>( s_2 = 0.72 )</td>
</tr>
</tbody>
</table>

16. **Hypothesis Test for Effects of Alcohol** Refer to the sample data in Exercise 15 and use a 0.05 significance level to test the claim that there is a difference between the treatment group and control group. If there is a significant difference, can we conclude that the treatment causes a decrease in visual and motor skills?

<table>
<thead>
<tr>
<th>Nonstress</th>
<th>Stress</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n_1 = 40 )</td>
<td>( n_2 = 40 )</td>
</tr>
<tr>
<td>( \bar{x}_1 = 53.3 )</td>
<td>( \bar{x}_2 = 45.3 )</td>
</tr>
<tr>
<td>( s_1 = 11.6 )</td>
<td>( s_2 = 13.2 )</td>
</tr>
</tbody>
</table>

17. **Hypothesis Test for Eyewitness Accuracy of Police** Does stress affect the recall ability of police eyewitnesses? This issue was studied in an experiment that tested eyewitness memory a week after a nonstressful interrogation of a cooperative suspect and a stressful interrogation of an uncooperative and belligerent suspect. The numbers of details recalled a week after the incident are summarized in the margin (based on data from “Eyewitness Memory of Police Trainees for Realistic Role Plays,” by Yuille et al., *Journal of Applied Psychology*, Vol. 79, No. 6). Use a 0.01 significance level to test the claim in the article that “stress decreased the amount recalled.”

18. **Confidence Interval for Eyewitness Accuracy of Police** Using the sample data from Exercise 17, construct a 98% confidence interval estimate of the difference between the two population means. Does the result support the claim in the article that “stress decreased the amount recalled”? Why or why not?
19. **Queen Mary Stowaways** Data Set 15 in Appendix B lists the ages of stowaway passengers on westbound and eastbound trips of the *Queen Mary*. When Excel is used with those two sets of ages, the results are as shown below. Is there a significant difference between the ages of stowaway passengers on westbound trips of the *Queen Mary* and the ages of stowaways on eastbound trips?

![Excel t-Test: Two-Sample Assuming Unequal Variances](image)

20. **Reading Levels** When a TI-83 Plus calculator is used with the Flesch reading ease scores for Tom Clancy’s *The Bear and the Dragon* and J. K. Rowling’s *Harry Potter and the Sorcerer’s Stone*, the accompanying results are obtained. (The sample data are listed in Data Set 14 in Appendix B.) Is there sufficient evidence to conclude that the mean Flesch reading ease score for Clancy is different than the mean for Rowling?

![TI-83 Plus 2-Sample t-Test](image)

21. **Tar and Cigarettes** Refer to the sample data listed below and use a 0.05 significance level to test the claim that the mean amount of tar in filtered king-size cigarettes is less than the mean amount of tar in nonfiltered king-size cigarettes. All measurements are in milligrams, and the data are from the Federal Trade Commission.

| Filtered | 16 | 15 | 16 | 14 | 16 | 1 | 16 | 18 | 10 | 14 | 12 | 11 | 14 | 13 | 13 | 16 | 16 | 8 | 16 | 11 |
|----------|----|----|----|----|----|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| Nonfiltered | 23 | 23 | 24 | 26 | 25 | 26 | 21 | 24 |

22. **Blanking Out on Tests** Many students have had the unpleasant experience of panicking on a test because the first question was exceptionally difficult. The arrangement of test items was studied for its effect on anxiety. The following scores are measures of “debilitating test anxiety,” which most of us call panic or blanking out (based on data from “Item Arrangement, Cognitive Entry Characteristics, Sex and Test Anxiety as Predictors of Achievement in Examination Performance,” by Klimko, Journal of Experimental Education, Vol. 52, No. 4.) Is there sufficient evidence to support the claim that the two populations of scores have the same mean? Is there sufficient evidence to support the claim that the arrangement of the test items has an effect on the score?

*continued*
Questions Arranged Questions Arranged
from Easy to Difficult from Difficult to Easy

<table>
<thead>
<tr>
<th>24.64</th>
<th>39.29</th>
<th>16.32</th>
<th>32.83</th>
<th>28.02</th>
<th>33.62</th>
<th>34.02</th>
<th>26.63</th>
<th>30.26</th>
</tr>
</thead>
<tbody>
<tr>
<td>33.31</td>
<td>20.60</td>
<td>21.13</td>
<td>26.69</td>
<td>28.90</td>
<td>35.91</td>
<td>26.68</td>
<td>29.49</td>
<td>35.32</td>
</tr>
<tr>
<td>26.43</td>
<td>24.23</td>
<td>7.10</td>
<td>32.86</td>
<td>21.06</td>
<td>27.24</td>
<td>32.34</td>
<td>29.34</td>
<td>33.53</td>
</tr>
<tr>
<td>28.89</td>
<td>28.71</td>
<td>31.73</td>
<td>30.02</td>
<td>21.96</td>
<td>27.62</td>
<td>42.91</td>
<td>30.20</td>
<td>32.54</td>
</tr>
<tr>
<td>25.49</td>
<td>38.81</td>
<td>27.85</td>
<td>32.34</td>
<td>21.96</td>
<td>27.62</td>
<td>42.91</td>
<td>30.20</td>
<td>32.54</td>
</tr>
</tbody>
</table>

23. BMI of Men and Women Refer to Data Set 1 in Appendix B and test the claim that the mean body mass index (BMI) of men is equal to the mean body mass index of women.

24. Marathon Runners Refer to Data Set 8 in Appendix B and test the claim that the mean age of a male runner in the New York City marathon is equal to the mean age of a female runner in that marathon.

In Exercises 25–28, assume that the two samples are independent simple random samples selected from normally distributed populations. Also assume that the population standard deviations are equal (\( \sigma_1 = \sigma_2 \)) so that the standard error of the differences between means is obtained by pooling the sample variances.

25. Confidence Interval with Pooling Do Exercise 7 with the additional assumption that \( \sigma_1 = \sigma_2 \). How are the results affected by this additional assumption?

26. Hypothesis Test with Pooling Do Exercise 8 with the additional assumption that \( \sigma_1 = \sigma_2 \). How are the results affected by this additional assumption?

27. Hypothesis Test with Pooling Do Exercise 9 with the additional assumption that \( \sigma_1 = \sigma_2 \). How are the results affected by this additional assumption?

28. Confidence Interval with Pooling Do Exercise 10 with the additional assumption that \( \sigma_1 = \sigma_2 \). How are the results affected by this additional assumption?

8-3 Beyond the Basics

29. Effects of an Outlier
   a. Refer to Exercise 19 and include an outlier consisting of a 90-year-old stowaway on a westbound crossing of the Queen Mary. Is the hypothesis test dramatically affected by the presence of the outlier?
   b. Refer to Exercise 19 and include an outlier consisting of a 5000-year-old stowaway on a westbound crossing of the Queen Mary. Why does the t test statistic decrease instead of increasing?

30. Calculating Degrees of Freedom How is the number of degrees of freedom for Exercises 13 and 14 affected if Formula 8-1 is used instead of selecting the smaller of \( n_1 - 1 \) and \( n_2 - 1 \)? If Formula 8-1 is used for the number of degrees of freedom instead of the smaller of \( n_1 - 1 \) and \( n_2 - 1 \), how are the \( P \)-value and the width of the confidence interval affected? In what sense is “\( df = \min(n_1 - 1, n_2 - 1) \)” a more conservative estimate of the number of degrees of freedom than the estimate obtained with Formula 8-1?