Chapter 9

1. Given sample sizes and number of successes to find the pooled estimate $\bar{p}$. Round your answer to the nearest thousandth.

$n_1 = 34, x_1 = 15; n_2 = 414, x_2 = 105$.  
Ans. $(x_1 + x_2)/(n_1 + n_2) = 120/448 \approx 0.268$

2. Assume that you plan to use confidence level $\alpha = 0.05$ to test the claim that $p_1 = p_2$. Use the following sample sizes and numbers of successes to find the $z$ test statistic for the hypothesis test.

In a vote on Clean Water bill, 44% of the 205 Democrats voted for the bill while 46% of the 230 Republicans voted for it.  
Ans. $z \approx -0.418 \ (\bar{p} \approx 0.483)$

3. The table shows the number satisfied in their work in a sample of working adults with a college education and in a sample of working adults without a college education. Find the critical value(s). Do the data provide sufficient evidence that a greater proportion of those with a college education are satisfied in their work? Assume that you plan to use a significance level of $\alpha = 0.05$ to test the claim that $p_1 > p_2$.

<table>
<thead>
<tr>
<th>College Education</th>
<th>No College Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number in sample</td>
<td>169</td>
</tr>
<tr>
<td>Number satisfied in their work</td>
<td>74</td>
</tr>
</tbody>
</table>

Ans. no ($p_1 = 74/169; p_2 = 68/162; \bar{p} = 142/331$)

claim: $p_1 > p_2$

Alternative: $p_1 \leq p_2$

Hypothesis test: $H_0: p_1 = p_2$

$H_1: p_1 > p_2$ (original claim) (right tail test)

$\alpha = 0.05$

Critical value: $z \approx 1.645$

Test statistic $z \approx 0.3329$

Fail to reject $H_0$

Final conclusion: there is not sufficient evidence to support the claim that $p_1 > p_2$.

4. Assume that you plan to use a significance level of $\alpha = 0.05$ to test the claim that $p_1 = p_2$. Use the given sample sizes and numbers of successes to find the P-value for the hypothesis test.

$n_1 = 100, x_1 = 41; n_2 = 140, x_2 = 35$.

Ans. $p_1 = 0.41; p_2 = 35/140 = 0.25; \bar{p} = 76/240 \approx 0.317$;

claim: $p_1 = p_2$

Alternative: $p_1 \neq p_2$

Hypothesis test: $H_0: p_1 = p_2$

$H_1: p_1 \neq p_2$ (Two tail test)

Test statistic $z \approx 2.626$;

P-value = $2\text{Normalcdf}(2.626, 10000, 0, 1) \approx 0.0086$

5. Use the traditional method to test the given hypothesis. Assume that all requirements are met.

In a random sample of 500 people aged 20-24, 22% were smokers. In a random sample of
450 people aged 25-29, 14% were smokers. Test the claim that the proportion of smokers in the two age groups is the same. Use the significance level of 0.01.

Ans. \( p_1 = 0.22, x_1 = 110; p_2 = 0.145, x_2 = 63; \quad \bar{p} = 173/950 \approx 0.182; \)

claim: \( p_1 = p_2 \)
Alternative: \( p_1 \neq p_2 \)
Hypothesis test: \( H_0: p_1 = p_2 \)
\[ H1: p_1 \neq p_2 \]
Test statistic: \[
t = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}} = \frac{0.22 - 0.14}{\sqrt{\frac{0.182(0.818)}{500} + \frac{0.182(0.818)}{450}}} \approx 3.19
\]
Critical value: \( z = \pm 2.575 \)

Reject \( H_0 \)
Final conclusion: there is sufficient evidence to warrant rejection of the claim that the proportion of smokers in the two age groups is the same.

6. The effectiveness of a new headache medicine is tested by measuring the amount of time before the headache is cured for patients who use medicine and another group of patients who use the placebo drug. Determine whether the samples are dependent or independent.

Ans. Independent

7. Test the indicated claim about the means of two populations. Assume that all requirements are met. Do not assume that \( \sigma_1 = \sigma_2 \). A researcher was interested in comparing the response times of two different cab companies. Company A and B were each called at 50 randomly selected times. The calls to company A were made independently of calls to company B. The response time for each call were recorded. The summary statistics were as follows:

<table>
<thead>
<tr>
<th></th>
<th>Company A</th>
<th>Company B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean response time</td>
<td>7.6 minutes</td>
<td>6.9 minutes</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>1.4 minutes</td>
<td>1.7 minutes</td>
</tr>
</tbody>
</table>

Use a 0.02 significance level to test the claim that the mean response time for company A is the same as the mean response time for company B. Use the P-value method.

Ans. \( \mu_1 = \mu_2 \)
Alternative: \( \mu_1 \neq \mu_2 \)
Hypothesis test: \( H_0: \mu_1 = \mu_2 \)
\[ H1: \mu_1 \neq \mu_2 \text{ (two tail test)} \]
\( \alpha = 0.02 \)
Critical value = 2.403
Test statistic \[
t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{7.6 - 6.9}{\sqrt{\frac{1.4^2}{50} + \frac{1.7^2}{50}}} \approx 2.248
\]
P-value = 0.029 (> 0.02)
Fail to reject \( H_0 \)
Final conclusion: there is not sufficient evidence to support the claim that \( \mu_1 = \mu_2 \).
8. A researcher was interested in comparing the amount of time spent watching television by women and by men. Independent simple random samples of 14 women and 17 men were selected. The summary data are as follows:

<table>
<thead>
<tr>
<th></th>
<th>Women</th>
<th>Men</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{x}_1$</td>
<td>12.9 hrs</td>
<td>$\bar{x}_2$ = 14.0 hrs</td>
</tr>
<tr>
<td>$s_1$</td>
<td>3.9 hrs</td>
<td>$s_2$ = 5.2 hrs</td>
</tr>
<tr>
<td>$n_1$</td>
<td>14</td>
<td>$n_2$ = 17</td>
</tr>
</tbody>
</table>

The following 98% confidence interval was obtained for $\mu_1 - \mu_2$,

$-6.028 < \mu_1 - \mu_2 < 3.828$. What does the confidence interval suggest about the population means?

**Ans.** Conclusion: since the CI contains 0, two population means might be equal.

9. The two data sets are dependent. Find $d$ to the nearest tenth.

<table>
<thead>
<tr>
<th>A</th>
<th>70</th>
<th>67</th>
<th>56</th>
<th>63</th>
<th>51</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>22</td>
<td>24</td>
<td>29</td>
<td>25</td>
<td>22</td>
</tr>
</tbody>
</table>

**Ans.** 37.0

10. Assume that you want to test the claim that the paired sample data come from a population for which the mean difference is $\mu_d = 0$. Compute the value of the t test statistic. Round the final answers to three decimal places.

<table>
<thead>
<tr>
<th>x</th>
<th>33</th>
<th>30</th>
<th>27</th>
<th>29</th>
<th>32</th>
<th>27</th>
<th>34</th>
<th>27</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>31</td>
<td>26</td>
<td>33</td>
<td>29</td>
<td>33</td>
<td>32</td>
<td>34</td>
<td>26</td>
</tr>
</tbody>
</table>

**Ans.** -0.523

11. Suppose you wish to test the claim that $\mu_d$ is different from 0. Given a sample of $n = 23$ and a significance level of $\alpha = 0.05$, what criterion would be used for rejecting the null hypothesis?

A. Reject null hypothesis if test statistic $> 1.717$.
B. Reject null hypothesis if test statistic $> 1.717$ or $< -1.717$.
C. Reject null hypothesis if test statistic $> 2.074$ or $< -2.074$.
D. Reject null hypothesis if test statistic $> 2.069$ or $< -2.069$.

**Ans.** C (critical values is $\pm 2.074$)

12. Assume that the population of paired differences is normally distributed. Ten different families are tested for the number of gallons of water a day they use before and after viewing a conservation video. Construct a 90% confidence interval for the mean of the differences.

<table>
<thead>
<tr>
<th>Before</th>
<th>33</th>
<th>33</th>
<th>38</th>
<th>33</th>
<th>35</th>
<th>35</th>
<th>40</th>
<th>40</th>
<th>40</th>
<th>31</th>
</tr>
</thead>
<tbody>
<tr>
<td>After</td>
<td>34</td>
<td>28</td>
<td>25</td>
<td>28</td>
<td>35</td>
<td>33</td>
<td>31</td>
<td>28</td>
<td>35</td>
<td>33</td>
</tr>
</tbody>
</table>

**Ans.** $1.8 < \mu_d < 7.8$ (mean = 4.8, $t_{0.05/2} = 1.833$, $E = 3.04$)

13. Assume that all requirements are met. The table below shows the weights of seven subjects before and after following a particular diet for two months.

<table>
<thead>
<tr>
<th>subject</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>before</td>
<td>180</td>
<td>188</td>
<td>172</td>
<td>193</td>
<td>195</td>
<td>168</td>
<td>158</td>
</tr>
<tr>
<td>after</td>
<td>173</td>
<td>179</td>
<td>179</td>
<td>198</td>
<td>181</td>
<td>170</td>
<td>146</td>
</tr>
</tbody>
</table>

Using a 0.01 level of significance, test the claim that the diet is effective in reducing weight.
**Ans.** Claim: \( \mu_d > 0 \)
Alternative: \( \mu_d \leq 0 \)
Hypothesis test: \( H_0: \mu_d = 0 \)
\[ H_1: \mu_d > 0 \text{ (original claim) (right tail)} \]
\( \alpha = 0.01 \)
Test statistic: \( t = \frac{\bar{d} - 0}{\frac{s_d}{\sqrt{n}}} = \frac{4 - 0}{8.52 / \sqrt{8}} \approx 1.327 \)
Critical Value = 2.998
Fail to reject \( H_0 \).
There is not sufficient evidence to support the claim that the diet is effective in reducing weight.

Chapter 10

1. In the regression equation \( \hat{y} = 31.6 + 10.9x \), \( x \) represents the number of years of study and \( y \) represents the grade on the test. Identify the predictor and the response variable.
   **Ans.** Number of years is the predictor variable, the grade on the test is the response variable.

2. Given that LCC \( r = -0.844 \), and \( n = 5 \). Find the critical values of \( r \) and determine whether or not the given \( r \) represents a significant linear correlation. Using \( \alpha = 0.05 \).
   **Ans.** critical values = \( \pm 0.878 \). Since \( | r | < 0.878 \), the given \( r \) does not represent a significant correlation.

3. The paired below consists of the test scores of 6 randomly selected students and the number of hours they studied for the test.

<table>
<thead>
<tr>
<th>hours</th>
<th>5</th>
<th>10</th>
<th>4</th>
<th>6</th>
<th>10</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>64</td>
<td>86</td>
<td>69</td>
<td>86</td>
<td>59</td>
<td>87</td>
</tr>
</tbody>
</table>

   Find the correlation coefficients.
   **Ans.** \( r = 0.224 \)

4. Describe the error in the stated conclusion. Given: there is no significant linear correlation between scores of a math test and scores on a verbal test. Conclusion: there is no relationship between scores on the math test and scores on the verbal test.
   **Ans.** there can be other relations (e.g. non-linear)

5. Use the given data to find the best predicted value of the response variable. Eight pairs of data yield \( r = 0.708 \) and the regression equation \( \hat{y} = 55.8 + 2.79x \). Also, \( \bar{y} = 71.125 \). What is the best predicted value of \( y \) for \( x = 5.7 \)?
   **Ans.** since \( r > 0.707 \) (at \( \alpha = 0.05 \)). The data are linearly correlated. So the best predicted value of \( y \) is \( 55.8 + 2.79(5.7) \approx 71.7 \).

6. Use the given data to find the equation of the regression line. Round the final values to three significant digits, if necessary.

<table>
<thead>
<tr>
<th>hours</th>
<th>0</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>8</td>
<td>2</td>
<td>6</td>
<td>9</td>
<td>12</td>
</tr>
</tbody>
</table>
Ans. $\hat{y} = 4.88 + 0.525x$

7. Using the same data as in the above problem. Find the residual at the sample point (3, 2)

Ans. Residual $\approx 2 - (4.88 + 0.525\cdot3) \approx -4.455$ (residual = observed – predicted)

8. Construct a scatterplot for the given data.

<table>
<thead>
<tr>
<th>hours</th>
<th>0</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>3</td>
<td>2</td>
<td>6</td>
<td>9</td>
<td>10</td>
</tr>
</tbody>
</table>

Provide an appropriate response.

9. The following residual plot is obtained after a regression equation is determined for a set of data. Does the residual plot suggest that the regression equation is a bad model? Why or why not?

Ans. No. Residual plot does not suggest that the regression equation is a bad model. (no pattern there)