ST 507  Reiland

PRACTICE PROBLEMS FOR FINAL EXAM

This material is covered in webassign homework assignments 10 through 12 and worksheets 13-18

Exam information: 3 hour time limit; materials allowed: calculator (no laptops or tablets),
1 8½x11 sheet of paper (2 sided) with notes, definitions, formulas, etc. Normal table, t-table, and chi-square table will be provided with the exam.
The questions on the exam will be multiple choice format.

Answers are at the end of the document.

Practice Problems

1. In a Risk Management Study on fires in compartmented fire-resistant buildings, the data below was generated. The data in the table give the number of victims who died trying to evacuate for a sample of 14 recent fires.

<table>
<thead>
<tr>
<th>FIRE</th>
<th>NUMBER OF VICTIMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Las Vegas Hilton (Las Vegas)</td>
<td>5</td>
</tr>
<tr>
<td>Inn on the Park (Toronto)</td>
<td>5</td>
</tr>
<tr>
<td>Westchase Hilton (Houston)</td>
<td>8</td>
</tr>
<tr>
<td>Holiday Inn (Cambridge, Ohio)</td>
<td>10</td>
</tr>
<tr>
<td>Conrad Hilton (Chicago)</td>
<td>4</td>
</tr>
<tr>
<td>Providence College (Providence)</td>
<td>8</td>
</tr>
<tr>
<td>Baptist Towers (Atlanta)</td>
<td>7</td>
</tr>
<tr>
<td>Howard Johnson (New Orleans)</td>
<td>5</td>
</tr>
<tr>
<td>Cornell University (Ithaca)</td>
<td>9</td>
</tr>
<tr>
<td>Westport Central Apartments (Kansias City, MO)</td>
<td>4</td>
</tr>
<tr>
<td>Orrington Hotel (Evanston IL)</td>
<td>0</td>
</tr>
<tr>
<td>Hartford Hospital (Hartford, CT)</td>
<td>16</td>
</tr>
<tr>
<td>Milford Plaza (New York)</td>
<td>0</td>
</tr>
<tr>
<td>MGM Grand (Las Vegas)</td>
<td>36</td>
</tr>
</tbody>
</table>

a. Construct a 98% confidence interval for the true mean number of victims per fire who die attempting to evacuate compartmented fire-resistant buildings.

b. Interpret the interval constructed in part a.

2. To investigate the possible link between fluoride content of drinking water and cancer, the cancer death rates (number of deaths per 100,000 population) from 1952-1969 in 20 selected U. S. cities - the ten largest fluoridated cities and the ten largest cities not fluoridated by 1969 - were recorded. These data were used to calculate for each city the annual rate of increase in cancer death rate over this 18 year period. The data are given below:
FLUORIDATED
City       Annual Increase in Cancer Death Rate
Chicago    1.0640
Philadelphia 1.4118
Baltimore  2.1115
Cleveland  1.9401
Washington 3.8772
Milwaukee  -1.4561
St. Louis  4.8359
San Francisco 1.8875
Pittsburgh 4.9464
Buffalo    1.4045

NONFLUORIDATED
City       Annual Increase in Cancer Death Rate
Los Angeles .8875
Boston    1.7358
New Orleans 1.0165
Seattle   .4923
Cincinnati 4.0155
Atlanta  -1.1744
Kansas City 2.8132
Columbus  1.7451
Newark    -.5676
Portland  2.4471

a. Construct a 95% confidence interval for the difference between the mean annual increases in cancer death rates for fluoridated and nonfluoridated cities. (Use 18 degrees of freedom).

3. Domino's Pizza in Big Rapids, Michigan, advertises that they deliver your pizza within 15 minutes of placing an order or it is free. A sample of 25 customers is selected at random. The average delivery time in the sample was 13 minutes with a sample standard deviation of 4 minutes.
a. Test to determine if we can infer at the 5% significance level that the population mean is less than 15 minutes.
b. Approximate the P-value for this test.

4. A marketing consultant was interested in estimating the mean weekly consumption of soft drinks among teenagers. A random sample of 61 teenagers were asked how many ounces of soft drink they consume daily. The sample mean and sample standard deviation are, respectively, 22.38 and 5.17. Estimate with 99% confidence the mean daily consumption of soft drinks by teenagers.

5. A random sample of 100 individuals was taken to determine the true percentage of people who smoke in a region of the eastern United States. Forty-six of them said “yes” when they were asked if they smoked. A 95% confidence interval for the true proportion of non-smokers is:
a. (.46, .54)  
b. (.36, .56)  
c. (.95, 1.0)  
d. (.44, .64)  
e. (.00, .95)

6. The credit manager of a department store would like to know what proportion of the credit-card customers take advantage of the store’s deferred payment plan each year. She would like to estimate this proportion within ±.10 at a 90% confidence level, but has no good idea about what this proportion might be. How many customers should she sample?
a. 271  
b. 41  
c. 17  
d. 68  
e. cannot be determined

7. When two competing teams are equally matched, the probability that each team wins any game is 0.5. The National Basketball Association (NBA) championship goes to the team that wins four games in a best-of-seven series. If the teams were evenly matched, the probability that the final series ends with one of the teams sweeping four straight games would be $2(0.5)^4 = 0.125$ [team A wins in 4 games with probability $(0.5)^4$; team B can also win in 4 games with the same probability, so the probability the series ends in 4 games is $2(0.5)^4$].

Similarly team A can win in 5 games if team A wins 3 of the first 4 games and then wins game 5. So team A wins in 5 games with probability $\binom{4}{3}(0.5)^3(0.5)^1(0.5) = 0.125$. But team B can also win in 5 games with the same probability, so the probability that the series ends in 5 games is 0.25. Similar probability calculations show that the probability is 0.3125 that the series lasts six games, and the probability is 0.3125 that the series lasts the full seven games. The table below shows the number of games it took to decide each of the last 57 NBA champs. Do you think the teams are usually equally matched? Give statistical evidence to support your conclusion.

<table>
<thead>
<tr>
<th>Length of series</th>
<th>4 games</th>
<th>5 games</th>
<th>6 games</th>
<th>7 games</th>
</tr>
</thead>
<tbody>
<tr>
<td>NBA finals</td>
<td>7</td>
<td>13</td>
<td>22</td>
<td>15</td>
</tr>
</tbody>
</table>
8. The president of a large university has been studying the relationship between male/female supervisory structures in his institution and the level of employees' job satisfaction. The results of a recent survey are shown in the table below. Conduct a test at the 5% significance level to determine whether the level of job satisfaction depends on the boss/employee gender relationship.

<table>
<thead>
<tr>
<th>Level of Satisfaction</th>
<th>Male/Female</th>
<th>Female/Male</th>
<th>Male/Male</th>
<th>Female/Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satisfied</td>
<td>60</td>
<td>15</td>
<td>50</td>
<td>15</td>
</tr>
<tr>
<td>Neutral</td>
<td>27</td>
<td>45</td>
<td>48</td>
<td>50</td>
</tr>
<tr>
<td>Dissatisfied</td>
<td>13</td>
<td>32</td>
<td>12</td>
<td>55</td>
</tr>
</tbody>
</table>

9. (True or false) In a hypothesis test, a P-value of 0.03 means that there is only probability 0.03 that the null hypothesis is true.

10. Suppose that 9 observations are drawn from a population whose distribution can be described by a normal model. The observations are:

15 9 13 11 8 12 11 7 10

Note that $\pi = 10.67$ and $s = 2.5$. You want to test whether the mean of the population from which this sample was taken is different from 12.

a. State the null and alternative hypotheses.
b. Compute the value of the test statistic.
c. Approximate the P-value.
d. What is your conclusion?

11. Sociologists are of the opinion that there has been a decrease in the difference in ages at first marriage for men and women since 1975. We want to examine data to determine if this decrease is significant. The following data summary and regression results were obtained, where the $x$ variable is year and the $y$ variable is the age difference (husband age – wife age) at first marriage.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Count</th>
<th>Mean</th>
<th>StDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year ($x$)</td>
<td>24</td>
<td>1986.5</td>
<td>7.071</td>
</tr>
<tr>
<td>husband-wife age ($y$)</td>
<td>24</td>
<td>2.3125</td>
<td>0.249</td>
</tr>
</tbody>
</table>
a. Interpret the value of the least squares slope $b_1$.

b. What is the value of the test statistic for testing $H_0 : \beta_1 = 0$?

c. For the hypothesis test $H_0 : \beta_1 = 0$ vs $H_a : \beta_1 < 0$, select the choice below that gives the correct $P$-value and correct conclusion.

i. The $P$-value is 0.68; do not reject $H_0$; there is no linear relationship since 1975 between year and age difference between husband and wife at first marriage.

ii. The $P$-value is 0.000152; reject $H_0$; there is evidence that since 1975 the age difference (husband age - wife age) has increased.

iii. The $P$-value is 0.0001275; reject $H_0$; there is evidence that since 1975 the age difference (husband age - wife age) has decreased.

iv. The $P$-value is 0.0001275; do not reject $H_0$; there is no linear relationship since 1975 between year and age difference between husband and wife at first marriage.

d. What is a 95% confidence interval for the slope?

12. Over 6 decades the Gallup Organization has periodically asked the following question:

If your party nominated a generally well-qualified person for president who happened to be a woman, would you vote for that person?

Below is a table showing the percentage answering “yes” and the year of the century (37 = 1937).

<table>
<thead>
<tr>
<th>% Yes</th>
<th>92</th>
<th>82</th>
<th>78</th>
<th>80</th>
<th>76</th>
<th>73</th>
<th>66</th>
<th>53</th>
<th>57</th>
<th>55</th>
<th>57</th>
<th>54</th>
<th>52</th>
<th>48</th>
<th>33</th>
<th>33</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>99</td>
<td>87</td>
<td>84</td>
<td>83</td>
<td>78</td>
<td>75</td>
<td>71</td>
<td>69</td>
<td>67</td>
<td>63</td>
<td>59</td>
<td>58</td>
<td>55</td>
<td>49</td>
<td>45</td>
<td>37</td>
</tr>
</tbody>
</table>

Summary statistics:

\[
\bar{x} = 67.44 \quad s_x = 16.7 \quad \bar{y} = 61.81 \quad s_y = 17.19 \quad r = .971
\]

a. Determine the values of the least squares intercept $b_0$ and least squares slope $b_1$ where $x$ is year and $y$ is the percentage who respond “yes”.

b. Use the least squares line to estimate the percentage of respondents that would say “yes” in 1997.

13. Which of the following statements is false?

a. The t distribution is symmetric about zero

b. The t distribution is more spread out than the standard normal distribution

c. As the degrees of freedom get smaller, the t-distribution's dispersion gets smaller

d. The t distribution is mound-shaped
14. The Student t distribution approaches the normal distribution as the:
   a. degrees of freedom increase
   b. degrees of freedom decrease
   c. sample size decreases
   d. population size increases

15. To test the hypotheses
   \[ H_0 : p = .06 \]
   \[ H_a : p < .06, \]
   a random sample of size 1000 revealed that the number of successes was 40. Compute the \( P \)-value and explain how to use it to test the hypothesis.

16. A professor claims that 70% of College of Business graduates earn more than $45,000 per year. In a random sample of 300 graduates, 195 earn more than $45,000. Perform a 2-tailed hypothesis test to test the professor's claim; compute the \( P \)-value for the test.

17. In the 1980's it was generally believed that congenital abnormalities affected about 5% of the nation's children. Some people believe that the increase in the number of chemicals in the environment has lead to an increase in the incidence of abnormalities. A recent study examined 408 children and found that 48 of them showed signs of an abnormality. Is this evidence that the risk has increased?
   i) Let \( p \) denote the proportion of children with genetic abnormalities. Choose the correct null and alternative hypotheses.
   A. vs  B.  vs
   C.    vs       D.     vs
   E.     vs       F.     vs

   ii) What is the value of the test statistic?
   iii) What is the \( P \)-value?
   iv) What is your conclusion?

18. A cell phone company wants to determine if the use of text messaging is independent of age. The following data has been collected from a random sample of customers.

<table>
<thead>
<tr>
<th></th>
<th>Regularly use text messaging</th>
<th>Do not regularly use text messaging</th>
</tr>
</thead>
<tbody>
<tr>
<td>Under 21</td>
<td>82</td>
<td>38</td>
</tr>
<tr>
<td>21-39</td>
<td>57</td>
<td>34</td>
</tr>
<tr>
<td>40 and over</td>
<td>6</td>
<td>83</td>
</tr>
</tbody>
</table>

18a. What is the expected value for the “under 21” and “regularly use text messaging” cell.

18b. To conduct a hypothesis test, the value of the critical value (use .05) is:
   i) 16.812     ii) 15.086     iii) 9.210   iv) 2.576   v) 2.33

18c. What contribution does the cell “21-39” and “Do not regularly use text messaging” make to the value of the test statistic?

18d. The value of the test statistic is 88.3. The appropriate conclusion is
   i) reject \( H_0 \) and conclude the variables have a curvilinear relationship;
   ii) reject \( H_0 \) and conclude the variables are not related, that is, conclude that the variables are independent.
   iii) do not reject \( H_0 \) and conclude the variables are independent.
   iv) reject \( H_0 \) and conclude the variables are related.
18e. Use the values of the standardized residuals \( \frac{\text{observed} - \text{expected}}{\text{expected}} \) to select the correct statement regarding the relationship between text messaging and age.

i) The "Under 21" age group uses text messaging less than the other age groups.

ii) For each age group the standardized residual has a negative value and a positive value, so text message and age are not related.

iii) The older the age group, the less that text messaging is used.

19. A chi-squared test for independence with 6 degrees of freedom results in a test statistic 13.61. Using the tables, the most accurate statement that can be made about the P-value for this test is that:

a. P-value > .10 

b. P-value > .05 

c. .05 < P-value < .10 

d. .025 < P-value < .05

20. A copy machine dealer has data on the number \( x \) of copy machines at each of 89 customer locations and the number \( y \) of service calls in a month at each location. Summary calculations give \( \bar{x} = 8.4 \), \( s_x = 2.1 \), \( \bar{y} = 14.2 \), \( s_y = 3.8 \), and \( r = .86 \). What is the slope of the least squares regression line of number of service calls on number of copiers?

21. In the setting of the previous problem, about what percent of the variation in number of service calls is explained by the linear relation between number of service calls and number of machines?

22. A study of 1,000 families gave the following results:

average height of husband = \( \bar{x} = 68 \) inches, \( s_x = 2.7 \) in.;

average height of wife = \( \bar{y} = 63 \) inches, \( s_y = 2.5 \) in.; \( r = .25 \).

Estimate the height of a wife when her husband is 72 inches tall.

a. 63 inches  
b. 72 inches  
c. 64 inches  
d. none of these  
e. need more information

Questions 23 and 24 refer to the following:

Data were collected to find the relationship between the labor (\( y \) in hours) required to produce lots of custom wood products and the size \( x \) of the lot. The following least squares regression equation was calculated from the data:

\[ \hat{y} = 13.7 + 1.7x. \]

23. What is the predicted hours of labor for a lot size of 55?

24. One of the original data points is (20, 50.3). What is the residual when the lot size is 20?

**SOLUTIONS**

1. a. From the 14 observations, \( \bar{x} = 8.36, s = 8.94; \) confidence coefficient .98 \( \Rightarrow \alpha = .02, \) from the \( t \)-table, \( t_{0.02} = 2.650, \)

\[ \bar{x} \pm t_{0.02,13}\left( \frac{s}{\sqrt{n}} \right) = 8.36 \pm 2.650\left( \frac{8.94}{\sqrt{14}} \right) = 8.36 \pm 6.33 = (2.03, 14.69) \]

b. “We are 98% confident that the interval (2.03, 14.69) encloses the true mean number of victims who die attempting to evacuate compartmented fire-resistant buildings.

2. a. \( \bar{x}_1 = 2.2573, \bar{x}_2 = 2.753, n_1 = 10; \bar{x}_2 = 1.3411, \bar{x}_2 = 2.429, n_2 = 10, t_{0.05,18} = 2.101, \)

\( (2.2573 - 1.3411) \pm 2.101\sqrt{\frac{2.2573}{10} + \frac{2.429}{10}} = .9162 \pm 1.5124 = ( -.5962, 2.4286). \)

3. a. \( H_0 : \mu = 15, H_a : \mu < 15; \) Rejection region: \( t < -t_{0.05,24} = -1.711; \) Test statistic: \( t = -2.50; \)

Conclusion: Reject the null hypothesis. Yes, we can conclude that the mean delivery time is less than 15 minutes.

3. b. Approximate P-value: the test statistic \( t = -2.5 \) is between \( -t_{0.024} = -2.492 \) and \( -t_{0.005} = -2.797. \) The P-value is between .005 and .01 (TI83: P-value = .0098)
4. \[ \bar{x} \pm t_{0.05, 60} \frac{s}{\sqrt{n}} = 22.38 \pm 2.66 \frac{5.37}{\sqrt{60}} = 22.38 \pm 1.76 = (20.62, 24.14) \]

5. d 6. d

7. \( H_0 \): teams are evenly matched  
   \( H_a \): teams are not evenly matched

The table below shows the observed values in each cell and the expected cell values in parentheses if the teams are evenly matched.

<table>
<thead>
<tr>
<th>Length of series</th>
<th>4 games</th>
<th>5 games</th>
<th>6 games</th>
<th>7 games</th>
</tr>
</thead>
<tbody>
<tr>
<td>NBA finals</td>
<td>7 (7.125)</td>
<td>13 (14.25)</td>
<td>22 (17.8125)</td>
<td>15 (17.8125)</td>
</tr>
</tbody>
</table>

Test statistic:
\[ X^2 = \frac{(7-7.125)^2}{7.125} + \frac{(13-14.25)^2}{14.25} + \frac{(22-17.8125)^2}{17.8125} + \frac{(15-17.8125)^2}{17.8125} = 1.54 \]

Rejection region: if the null hypothesis is true, the test statistic \( X^2 \) has a chi-square distribution with \((k-1) = (4-1) = 3 \) dfs. If \( \alpha = .05 \), the RR is \( X^2 > 7.81 \). Note that \( P-value = 0.67 \)

Conclusion: do not reject \( H_0 \). There is no evidence that the NBA championship series are inconsistent with the conjecture that the teams are evenly matched.

8. (chap 26)

Expected cell counts are in parentheses:

<table>
<thead>
<tr>
<th>Level of Satisfaction</th>
<th>Male/Female</th>
<th>Female/Male</th>
<th>Male/Male</th>
<th>Female/Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satisfied</td>
<td>60 (33.175)</td>
<td>15 (30.521)</td>
<td>50 (36.493)</td>
<td>15 (39.81)</td>
</tr>
<tr>
<td>Neutral</td>
<td>27 (40.284)</td>
<td>45 (37.062)</td>
<td>48 (44.313)</td>
<td>50 (48.341)</td>
</tr>
<tr>
<td>Dissatisfied</td>
<td>13 (26.54)</td>
<td>32 (24.417)</td>
<td>12 (29.194)</td>
<td>55 (31.848)</td>
</tr>
</tbody>
</table>

\( H_0 \): Boss/employee relationship and job satisfaction are independent  
\( H_a \): Boss/employee relationship and job satisfaction are dependent

Test statistic: \( \chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}} = 92.709 \); degrees of freedom \( = (3-1)(4-1) = 6 \)

\( P-value = .000 \). Conclusion: reject \( H_0 \) and conclude that boss/employee relationship and job satisfaction are related.

9. false

10. a. \( H_0 : \mu = 12, H_a : \mu \neq 12 \); b. \( \bar{x} = 10.67, s = 2.5, t = \frac{10.67-12}{2.5} = -1.596 \).

   c. Use \( t \) distribution with 8 degrees of freedom; from the \( t \)-table \( P(t < -1.596) + P(t > 1.596) \) is between 0.1 and 0.2. d. Do not reject \( H_0 \); there is no evidence that \( \mu \) differs significantly from 12.

11. a. \( b_1 = -0.024 \) (approximately); this means that each year since 1975 the average difference (husband age − wife age) has decreased by .024

   b. \( t = \frac{b_1}{s_{\text{estimate}}} = \frac{-0.023956}{0.085363} = -2.786 \)

   c. iii. The \( P \)-value given in the Excel output is always for a 2-tail test; since we are conducting a 1-tail test \( H_a : \beta_1 < 0 \), the \( P \)-value is \( \frac{0.000025}{2} = 0.0000125 \)

   d. \((-0.0353697, -0.01254335)\) from the output; notice that the interval is entirely negative.

12. a. \( b_1 = r_{xy} = \frac{971.1719}{17} = .99949; b_0 = \bar{y} - b_1 \bar{x} = 61.81 - .99949(67.44) = -5.59358 \)

   b. \( \hat{y}_{97} = -5.59358 + .99949(97) = 91.36 \)

13. c 14. a

15. \( \hat{p} = \frac{40}{1000} = .04 \); test statistic \( z = \frac{.04-.06}{\sqrt{.00008}} = -2.67 \); \( P-value = P(z < -2.67) = .0038 \).

   Since the \( P \)-value is less than .05, reject \( H_0 \) and conclude that \( p < .06 \).

16. \( H_0 : p = .7, H_a : p \neq .7 \);

   \( \hat{p} = \frac{105}{300} = .65; SD(\hat{p}) = \sqrt{\frac{(1)(.3)}{300}} = .02646 \);
test statistic $z = -1.89$; $P-value = 2P(z > | -1.89|) = 2(.0294) = .0588$

do not reject $H_0$; there is no evidence to support the claim that the proportion of graduates who earn more than $45,000 differs significantly from .70

17. i) D  
   ii) $\hat{p} = \frac{48}{408} = 0.1176$; $SD(\hat{p}) = \sqrt{\frac{(0.05)(1-0.05)}{408}} = \sqrt{\frac{0.0475}{408}} = 0.0108$;
   
   $z = \frac{0.1176 - 0.05}{0.0108} = 6.26$;
   
   iii) $P-value = P(z > 6.26) \approx 0$; iv) Since the $P-value$ is almost zero, there is strong evidence against the null hypothesis $H_0 : p = 0.05$; we reject $H_0 : p = 0.05$ and conclude that the risk has increased.

18. |             | Regularly use text messaging | Do not regularly use text messaging |
      | Under 21    | 82 (58)                  | 38 (62)                  |
      | 21-39       | 57 (43.983)              | 34 (47.017)              |
      | 40 and over | 6 (43.017)               | 83 (45.983)              |

18a. expected count: $\frac{row total \times column total}{grand total} = 58.$

18b. iii) 5.991 ($df = (3-1) \times (2-1)$)

18c. $\frac{(observed - expected)^2}{expected} = \frac{(34-47.017)^2}{47.017} = 3.604$

18d. iv) reject $H_0$ and conclude the variables are related.

18e. iii) The older the age group, the less that text messaging is used. (see table below)

<table>
<thead>
<tr>
<th></th>
<th>Regularly use text messaging</th>
<th>Do not regularly use text messaging</th>
</tr>
</thead>
<tbody>
<tr>
<td>Under 21</td>
<td>3.15</td>
<td>-3.05</td>
</tr>
<tr>
<td>21-39</td>
<td>1.96</td>
<td>-1.90</td>
</tr>
<tr>
<td>40 and over</td>
<td>-5.64</td>
<td>5.46</td>
</tr>
</tbody>
</table>

19. d 20. $b = r \left( \frac{S_y}{S_x} \right) = .86 \left( \frac{3.8}{27} \right) = 1.56.$  

21. $r^2 = (.86)^2 = .74$

22. c. $b_1 = r* \left( \frac{S_y}{S_x} \right) = .25* \left( \frac{3.8}{27} \right) = .231$; $b_0 = \bar{y} - b_1\bar{x} = 63 - .231*68 = 47.292$; 

$\hat{y} = 47.292 + .231*72 = 63.924 \approx 64.$

23. 107.2  
24. observed $y$ - predicted $y = 2.6$ hours.