Material covered on exam:

✓ text chapters 19 - 27
✓ webassign homework #9 through #12

Needed for exam:

✓ 8 ½" x 11" sheet of notes, formulas (handwritten or typed, 2-sided); calculator
  (I will supply needed tables)

1. A simple linear regression was used to predict the score \( y \) on a final exam from the score \( x \) on the first exam. The slope of the least squares regression line is .75. The standard error of the slope is .11 and the sample size is 200. A 90% confidence interval for the true slope is (though the \( df \) is \( n - 2 \), use 200 \( df \) in the \( t \)-table)

   a. .64 to .86  
   b. .53 to .97  
   c. .57 to .93  
   d. .64 to .86  
   e. none of the above

2. Refer to the previous problem. To test the null hypothesis that the slope is zero versus the one-sided alternative that the slope is positive, we use the test statistic

   a. 6.82  
   b. .05  
   c. .15  
   d. .95  
   e. .75

3. The failure rate (or hazard rate) \( h(t) \) of a system is related to the conditional probability that the system will fail in the next instant of time given that it has survived to time \( t \). The hazard rate of a particular brand of personal computer is thought to be a power function of its age \( t \), so that

   \[ h(t) = c t^d, \]

where \( c \) and \( d \) are parameters with unknown values that need to be estimated. To convert the above expression to a linear equation involving \( c \) and \( d \) we take the log (base 10) of each side to obtain

   \[ \log h(t) = \log c + d \log t, \]

or

   \[ y = \beta_0 + \beta_1 x \]

where \( y = \log h(t) \), \( \beta_0 = \log c \), \( \beta_1 = d \), and \( x = \log t \). A large number of computers were put on test for 10 hours to obtain an estimate of the hazard rate for each of the ten hours. The results are shown below:

<table>
<thead>
<tr>
<th>( t ) (hours)</th>
<th>( x_i = \log_{10} t_i )</th>
<th>( y_i = \log_{10} h_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0000</td>
<td>0.02119</td>
</tr>
<tr>
<td>2</td>
<td>0.3010</td>
<td>0.59600</td>
</tr>
<tr>
<td>3</td>
<td>0.4771</td>
<td>0.91381</td>
</tr>
<tr>
<td>4</td>
<td>0.6021</td>
<td>1.24551</td>
</tr>
<tr>
<td>5</td>
<td>0.6990</td>
<td>1.39794</td>
</tr>
<tr>
<td>6</td>
<td>0.7782</td>
<td>1.57978</td>
</tr>
<tr>
<td>7</td>
<td>0.8451</td>
<td>1.69020</td>
</tr>
<tr>
<td>8</td>
<td>0.9031</td>
<td>1.77085</td>
</tr>
<tr>
<td>9</td>
<td>0.9542</td>
<td>1.92942</td>
</tr>
<tr>
<td>10</td>
<td>1.0000</td>
<td>1.98900</td>
</tr>
</tbody>
</table>

The following straight line model was proposed:

   \[ y = \beta_0 + \beta_1 x + \epsilon. \]

Summary statistics for the above data are:

\[ n = 10 \quad \bar{x} = .65598, \quad s_x = .31835 \quad \bar{y} = 1.31337 \quad s_y = .63369 \]

\[ r = .9992 \]

\[ \sum x_i = 6.5598 \quad \sum x_i^2 = 5.2152 \quad \sum y_i = 13.1337 \quad \sum x_i y_i = 10.4295 \]
a. Calculate $b_0$, $b_1$, the least squares estimates of the model parameters $\beta_0$ and $\beta_1$.

b. Calculate an estimate of $\sigma_e$, the standard deviation of the random error ($\epsilon$) component of the model.

c. Calculate the coefficient of determination and interpret its value.

For the remainder of the problem, you may assume that the following values have been calculated:

$b_0 = .0087 \quad b_1 = 1.989 \quad s_e = .02724$

d. Set up the calculation of a 95% confidence interval for $\beta_1$. Suppose the confidence interval is $(1.9226, 2.0554)$. Does it appear that $x (= \log t)$ is useful for predicting $y (= \log h(t))$?

e. Set up the calculation of a 95% prediction interval for $y$ when $x = .5$

f. What assumptions are required for the valid use of the confidence interval procedures in parts (d) and (e)?

4. 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?

5. 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?

6. Each of the following statements contains a blunder. In each case explain what is wrong.
   a. "There is a high correlation between the sex of American workers and their income."
   b. "We found a high correlation ($r = 1.09$) between students' ratings of faculty teaching and ratings made by other faculty members."
   c. "The correlation between planting rate and yield of corn was found to be $r = .23$ bushel."

7. 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 1.5 standard deviations above the mean husband height.
   a. 63 inches  b. 72 inches  c. 64 inches  d. none of these  e. need more information

8. Shown below is a scatterplot with the corresponding least squares line.

Choose the residual plot that corresponds to the above scatterplot and least squares line.

a. I  b. II  c. III  d. IV  e. none
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. 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 (Kansas 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>

*Source: Macdonald, J. N. “Is Evacuation a Fatal Flaw in Fire-Fighting Philosophy?”*  

\[ \sum_{i=1}^{14} x_i = 117; \sum_{i=1}^{14} (x_i - \bar{x})^2 = 1039.214 \]

a. We would like to calculate a confidence interval for the mean number $\mu$ of victims per fire. State the assumption, in terms of the problem, that is required for the confidence interval technique to be valid.
b. Construct a 98% confidence interval for the true mean number of victims per fire who die attempting to evacuate compartmented fire-resistant buildings.

c. Interpret the interval constructed in part b.

11. 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

12. 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

13. The statistic $\frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$ has the student t distribution only if the sample is drawn from:
   a. Student t distribution
   b. a normal distribution
   c. a negatively skewed distribution (skewed to the left)
   d. a positively skewed distribution (skewed to the right)

14. 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. Perform a hypothesis test to determine if we can conclude that the population mean is less than 15 minutes.
   b. What is the required condition of the technique used in part (a)?
   c. Approximate the P-value for this test.

15. A marketing consultant was interested in estimating the mean daily consumption of soft drinks among teenagers. A random sample of 61 teenagers were asked how many ounces of soft drink they consume daily. The sum of the observations and the sum of the squared deviations of the observations from the overall mean $\bar{x}$ are shown below.

$$\sum_{i=1}^{61} x_i = 1365 \quad \text{and} \quad \sum_{i=1}^{61} (x_i - \bar{x})^2 = 1605.328$$

Estimate with a 99% confidence interval the mean daily consumption of soft drinks by teenagers.

16. As the sample size $n$ increases, when the confidence level is held fixed the width of the confidence interval for the population mean tends to:
   a. increase
   b. decrease
   c. stay the same

17. A random sample of 49 data values had sample standard deviation $s = 8.2$. Suppose the confidence interval for $\mu$ calculated from the data was $63 < \mu < 69$. What is the sample mean $\bar{x}$?
18. A confidence level of 99% can correctly be interpreted to mean that
   a. 99% of the time in repeated sampling, intervals calculated using an appropriate formula will
       contain the sample value.
   b. 99% of the time in repeated sampling, intervals calculated using an appropriate formula will
       contain the relevant population parameter.
   c. 99% of the time in repeated sampling, intervals calculated using an appropriate formula will
       contain the sample value as the midpoint of the interval.
   d. 99% of the time in repeated sampling, intervals calculated using an appropriate formula will
       contain the sample mean as their midpoints.

19. Suppose that 9 observations are drawn from a population whose distribution can be approximated by a
    normal model. The observations are:
    
    15  9  13  11  8  12  11  7  10

    Note that \( \bar{x} = 10.67 \) 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?

20. Recently there have been campaigns encouraging people to save energy by carpooling to work. Some
    cities have created “carpool only” traffic lanes (i.e. only cars with 2 or more passengers can use these
    lanes). In order to evaluate the effectiveness of carpool only lanes, toll booth personnel in one city
    monitor 2,000 randomly selected cars in 2005 before the carpool lanes were established and 1,500 cars
    in 2010 after the lanes were established. The results are shown below, where \( x_1 \) (\( x_2 \)) is the number of
    cars with 2 or more passengers in the data for 2005 (2010). Use a 95% confidence interval to
    determine whether the data indicate that the fraction of cars with carpool riders has increased over this
    period.
    
    2005: 2,000, 576; 2010: 1,500, 652.

21. The lower limit of a confidence interval at the 95% level of confidence for the population proportion
    if a sample of size 200 had 40 successes is:
    a. .2554  b. .1446  c. .2465  d. .1535

22. Perform the hypothesis test shown below,
    \[ H_0 : p = .05 \]
    \[ H_a : p < .05 \]
    given that a random sample of size 1000 revealed that the number of successes was 40. Compute
    the P-value and use it to make a conclusion concerning the hypothesis test.

23. 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 hypothesis test to evaluate
    the professor's claim.

24. 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 1982-1999 in 20 selected U.S. cities - the
    ten largest fluoridated cities and the ten largest cities not fluoridated by 1999 - 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          NONFLUORIDATED

<table>
<thead>
<tr>
<th>City</th>
<th>Annual Increase in Cancer Death Rate</th>
<th>City</th>
<th>Annual Increase in Cancer Death Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chicago</td>
<td>1.0640</td>
<td>Los Angeles</td>
<td>.8875</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>1.4118</td>
<td>Boston</td>
<td>1.7358</td>
</tr>
<tr>
<td>Baltimore</td>
<td>2.1115</td>
<td>New Orleans</td>
<td>1.0165</td>
</tr>
<tr>
<td>Cleveland</td>
<td>1.9401</td>
<td>Seattle</td>
<td>.4923</td>
</tr>
<tr>
<td>Washington</td>
<td>3.8772</td>
<td>Cincinnati</td>
<td>4.0155</td>
</tr>
<tr>
<td>Milwaukee</td>
<td>-.4561</td>
<td>Atlanta</td>
<td>-1.1744</td>
</tr>
<tr>
<td>St. Louis</td>
<td>4.8359</td>
<td>Kansas City</td>
<td>2.8132</td>
</tr>
<tr>
<td>San Francisco</td>
<td>1.8875</td>
<td>Columbus</td>
<td>1.7451</td>
</tr>
<tr>
<td>Pittsburgh</td>
<td>4.4964</td>
<td>Newark</td>
<td>-.5676</td>
</tr>
<tr>
<td>Buffalo</td>
<td>1.4045</td>
<td>Portland</td>
<td>2.4471</td>
</tr>
</tbody>
</table>

a. Construct a 95% confidence interval for the difference between the mean annual increases in cancer death rates for fluoridated and nonfluoridated cities.

b. Let \( \mu_1 \) be the mean annual increase in the cancer death rate for fluoridated cities and let \( \mu_2 \) be the mean annual increase in the cancer death rate for nonfluoridated cities. Perform a hypothesis test to investigate if the mean annual increase in the cancer death rate for fluoridated cities is greater than the mean annual increase in the cancer death rate for nonfluoridated cities.

25. The president of a large university has been studying the relationship between male/female supervisory structures in her 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>

26. 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 2 teams are evenly matched, the probability that the series ends with one of the teams winning the first four 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 \( \binom{4}{3}(0.5)^3(0.5)(0.5) = 0.125 \). 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>

27. 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>

**Regression Statistics**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple R</td>
<td>0.680298874</td>
</tr>
<tr>
<td>R Square</td>
<td>0.462753496</td>
</tr>
<tr>
<td>Adj. R Square</td>
<td>0.43833291</td>
</tr>
<tr>
<td>Standard Error</td>
<td>0.18662649</td>
</tr>
<tr>
<td>Observations</td>
<td>24</td>
</tr>
</tbody>
</table>

**ANOVA**

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>$F$</th>
<th>Signif $F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>0.660002174</td>
<td>0.660002</td>
<td>10.94955</td>
<td>0.000254659</td>
</tr>
<tr>
<td>Residual</td>
<td>22</td>
<td>0.766247926</td>
<td>0.034829</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>23</td>
<td>1.42625</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Coefficients**

<table>
<thead>
<tr>
<th></th>
<th>Standard Error</th>
<th>$t$ Stat</th>
<th>$P$-value</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>49.90213043</td>
<td>10.9324015</td>
<td>4.564608</td>
<td>0.000152</td>
<td>72.22571749</td>
</tr>
<tr>
<td>Year</td>
<td>-0.023953622</td>
<td>-0.05503315</td>
<td>-4.35311</td>
<td>0.000255</td>
<td>-0.0353697</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 $H_0 : \beta_1 = 0$; 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 $H_0 : \beta_1 = 0$; there is evidence that since 1975 the age difference (husband age – wife age) has increased.
   iii. The $P$-value is 0.0001275; reject $H_0 : \beta_1 = 0$; 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 $H_0 : \beta_1 = 0$; 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?
e. Find a 95% confidence interval for the mean difference (husband age – wife age) at first marriage in 1998.
f. Find a 95% prediction interval for the difference (husband age – wife age) for a particular couple getting married for the first time in 1998.

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

**Summary statistics:**
Determine the estimates $b_0$ and $b_1$ of the parameters $\beta_0$ and $\beta_1$ in the linear model $y = \beta_0 + \beta_1 x + \epsilon$, 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.

c. Determine the estimate $s_e$ of the standard deviation $\sigma$ of the error component $\epsilon$ (note that the sum of squares of residuals $\sum_{i=1}^{10} (y_i - \hat{y}_i)^2 = 255.748$).

d. Calculate a 95% confidence interval for the slope $\beta_1$. (Note that $SE(b_1) = \frac{s_e}{\sqrt{n-1}s_\bar{x}}$)

e. Conduct an appropriate hypothesis test (use $\alpha = .05$) to determine if the year of the century is useful for predicting the percentage of respondents that would answer “yes” to the above question. State the hypotheses, find the value of the test statistic, and state your conclusion based on the $P$-value or the rejection region.

**SOLUTIONS**

1. $b_1 \pm t^*SE(b_1)$

2. $t = \frac{b_1}{SE(b_1)}$

3. $b_1 = .9992(\frac{.6339}{.3158}) = 1.98895; b_0 = 1.31337 - b_1.65598 = .008659$;

b. $s_e = \sqrt{SSE/(n-2)}$ where $SSE = \sum_{i=1}^{10} y_i^2 - b_0 \sum_{i=1}^{10} y_i - b_1 \sum_{i=1}^{10} x_i y_i = .005944$; so

$c. r^2 = .9992^2 = .9984; \text{if } r^2 = .9984, \text{then 99.84% of the variation in log } h(t) \text{ about its mean is explained by the linear relationship with log } t.$

d. $1.989 \pm 2.306(0.2724)/\left(\sqrt{9*.31835}\right)$. Since the confidence interval for the slope is entirely positive, it appears that the slope $\beta_1$ is not 0 and that $x (= \log t)$ is useful for predicting $y (= \log h(t))$.

e. $\hat{y} \pm t^*SE(\hat{y}_x)$; $\hat{y}_x = .0087 + 1.989*.5 = 1.0032; \ t^* = 2.3060 (8 df);$ $SE^2(b_1) = \left(\frac{s}{\sqrt{10-1}s_\bar{x}}\right)^2 = (\frac{.0275}{.3158})^2 = .0288^2 = .00083$ $SE(\hat{y}_x) = \sqrt{SE^2(b_1) \times (x_0 - \bar{x})^2 + \frac{s_0^2}{n} + s_e^2} = \sqrt{.00083 \times (.5 - .65598)^2 + .00076^2 + .00076^2} = .0293$ $\hat{y} \pm t^*SE(\hat{y}_x) = 1.0032 \pm 2.3060(.0293) = 1.0032 \pm .0676$

f. $\epsilon \sim N(0, \sigma)$ for all $x$ and are independent across observations.

4. $b_1 = r\left(\frac{\bar{y}}{\bar{x}}\right) = .86(\frac{13}{22}) = 1.56$. 5. $r^2 = (.86)^2 = .74$

6. a. The correlation we are studying measures the linear relationship between 2 quantitative variables; sex is a categorical variable.

b. $-1 \leq r \leq 1$ is violated.

c. $r$ has no units.

7. The husband's height 1.5 $x$ standard deviations above the mean husband height of 68 inches. The wife's height is predicted to be above average by .25(1.5) = .4 $y$ standard deviations, or $A \times 2.5$ inches = 1 inch. (Recall $b = r(s_y/s_x)$)

8. c. III 9. false

10. a. The distribution of the population of the number of victims who attempt to evacuate fires can be approximated by a normal model.

b. From the 14 observations, $\bar{x} = 8.36, s = 8.94; \text{confidence coefficient } .98 \Rightarrow \text{from the } t\text{-table}, \ t^* = 2.6503 (13 df)$ $\bar{x} \pm t^*\left(\frac{s}{\sqrt{n}}\right) = 8.36 \pm 2.6503\left(\frac{8.94}{\sqrt{14}}\right) = 8.36 \pm 6.33 = (2.03, 14.69)$
c. “We are 98% confident that the interval (2.03, 14.69) contains the true mean number of victims who die attempting to evacuate compartmented fire-resistant buildings.

11. a. $H_0: \mu = 15$, $H_a: \mu < 15$; Test statistic: $t = -2.50$; b. The distribution of pizza delivery times can be approximated by a normal model.

15. c. Approximate P-value: the test statistic $t = -2.50$ is between $-2.4922$ and $-2.7970$. The P-value is between $.005$ and $.01$ (TI83: $P = .0098$)

16. $s = \sqrt{\frac{1605.328}{60}} = 5.17; \overline{x} \pm t_{0.025, 61} = 22.38 \pm 2.6603(5.17) = 22.38 \pm 1.76 = (20.62, 24.14)$.

17. b. decrease 18. 66 19. b

20. a. $H_0: \mu = 12$, $H_a: \mu \neq 12$; b. $\overline{x} = 10.67, s = 2.5; t = \frac{10.67 - 12}{2.5} = -1.596$.

21. $\hat{p}_1 = \frac{576}{2000} = .288; \hat{p}_2 = \frac{652}{1500} = .435$; the confidence interval for $p_1 - p_2$ is

$$(.288 - .435) \pm 1.96 \times \sqrt{\frac{(.288 \times .712) + (.435 \times .565)}{2000}} = -.147 \pm .032 \Rightarrow (-.179, -.115)$$

Conclusion: Conclude that the proportion of all cars with carpool riders has increased over the 2005-2010 period.

22. b 33. Test statistic $z = -1.451$; P-value $P(z < -1.451) = .0734$. Since the P-value is greater than $\alpha = .05$, do not reject $H_0$; there is no evidence that $\mu$ differs significantly from 12.

23. $H_0 : p = .7, H_a : p \neq .7; \hat{p} = \frac{105}{150} = .65$; test statistic $z = \frac{.65 - .7}{\sqrt{.10}} = -1.89$; P-value $P(z < -1.89) + P(z > 1.89) = .0588$. Since P-value > .05, do not reject the null hypothesis; there is no evidence that $p$ differs significantly from .7

24. $\overline{x}_1 = 2.2573, s_1^2 = 2.753, n_1 = 10; \overline{x}_2 = 1.3411, s_2^2 = 2.429, n_2 = 10$;

a. Use degrees of freedom $= \min(n_1 - 1, n_2 - 1) = \min(9, 9) = 9; t_{0.025,9} = 2.2622$;

$b = 2.573 - 1.3411 \pm 2.262 \sqrt{\frac{2.753}{10} + \frac{2.429}{10}} = .9162 \pm 1.628 = (-.718, 2.5442)$. Note that this interval will differ from the interval given on the TI-83 and TI-84 since the calculators use different degrees of freedom.

b. $H_0: \mu_1 - \mu_2 = 0$, $H_a: \mu_1 - \mu_2 > 0$.

Test statistic: $t = \frac{2.2573 - 1.3411}{\sqrt{\frac{2.753}{10} + \frac{2.429}{10}}} = 1.27$; P-value $= P(t > 1.27); \text{use degrees of freedom} = \min(n_1 - 1, n_2 - 1) = \min(9, 9) = 9$ to approximate the P-value; from the t-table, P-value $= P(t > 1.27)$ is between 0.1 and 0.15. Therefore, do not reject the null hypothesis; there is no evidence that the mean annual increase in cancer death rate is significantly greater for fluoridated cities than for nonfluoridated cities.

25. Expected cell counts are in parentheses:

<table>
<thead>
<tr>
<th>Level of Satisfaction</th>
<th>Boss/Female</th>
<th>Female/Male</th>
<th>Male/Female</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>
<td></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>
<td></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>
<td></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: $\chi^2 = \sum_{1}^{12} \frac{(\text{observed} - \text{expected})^2}{\text{expected}} = 92.709$; degrees of freedom $= (3 - 1) \times (4 - 1) = 6$

Cutoff value from chi-square table is 12.592. Conclusion: reject $H_0$ and conclude that boss/employee relationship and job satisfaction are related.

26. $H_0$: teams are evenly matched
ST 511 Practice Problems for Final Exam

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

degrees of freedom = (4 - 1) = 3; cutoff value from chi-square table is 7.815.

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.

27. 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. \( t = \frac{b_1}{SE(b_1)} = -0.023956 \cdot \frac{0.00363}{0.00363} = -4.35311 \)

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 \( \frac{0.0001275}{2} = 0.00006375 \)

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

e. use \( \hat{y}_{1998} \pm t_{n-2}SE(\hat{y}_{1998}) \); for the calculations below, note that from the output we have \( s_e = 0.18662649; \)

\( \hat{y}_{1998} = 49.90213043 - 0.023956522(1998) = 2.037 \)

\( SE(\hat{\mu}_{1998}) = \sqrt{SE^2(\hat{y}_{1998}) \times (x_n - \bar{x})^2 + \frac{s^2}{n}} = \sqrt{(0.05503315)^2 \times (1998 - 1986.5)^2 + \frac{0.18662649^2}{24}} = .7386889; \)

\( t_{n-2}SE(\hat{\mu}_{1998}) = 2.074(.7386889) = .153204; \)

so \( \hat{y}_{1998} \pm t_{n-2}SE(\hat{\mu}_{1998}) = 2.037 \pm .153204 \Rightarrow (1.883796, 2.190204) \)

f. use \( \hat{y}_{1998} \pm t_{n-2}SE(\hat{y}_{1998}) \); for the calculations below, note that from the output we have \( s_e = 0.18662649; \)

\( \hat{y}_{1998} = 49.90213043 - 0.023956522(1998) = 2.037; \)

\( SE(\hat{y}_{1998}) = \sqrt{SE^2(\hat{y}_{1998}) \times (x_n - \bar{x})^2 + \frac{s^2}{n}} = \sqrt{(0.05503315)^2 \times (1998 - 1986.5)^2 + \frac{0.18662649^2}{24}} = .41628; \)

\( t_{n-2}SE(\hat{y}_{1998}) = 2.074(.200714) = .41628; \)

so \( \hat{y}_{1998} \pm t_{n-2}SE(\hat{y}_{1998}) = 2.037 \pm .41628 \Rightarrow (1.62072, 2.45328) \)

28. a. \( b_1 = \frac{\alpha}{\sqrt{n}} = .97147; b_0 = \bar{y} - b_1\bar{x} = 61.81 - .99949(67.44) = -5.9358; \)

b. \( \hat{y}_{97} = -5.9358 + .99949(97) = 91.36 \)

c. \( s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{255.748}{14}} = 4.274; \)

d. \( SE(b_1) = \frac{s_e}{\sqrt{n-1}s_x} = \frac{4.274}{\sqrt{15} \cdot 16.7} = .661; \)

confidence interval is \( b_1 \pm t_{n-2}SE(b_1) = .99949 \pm 2.145(.0661) = .99949 \pm .14178 \Rightarrow (.85771, 1.14127) \)

e. \( H_0: \beta_1 = 0 \) vs \( H_a: \beta_1 \neq 0; \)

test statistic \( t = \frac{b_1}{SE(b_1)} = \frac{.99949}{.0661} = 15.12; \)

for \( \alpha = .05 \), the rejection region is \( t > 2.145 \) and \( t < -2.145 \); the \( P \)-value is 0 to nine decimal places.

Conclusion: since the test statistic is in the rejection region, reject \( H_0: \beta_1 = 0 \) and conclude that year is useful for predicting the percentage of respondents that will answer “yes” to the question.