Name:________________________

Write your answers to the first five questions on the attached pages, in the spaces provided. Circle the choice which best answers questions 6-15. Do not write anything else on this page (besides your name and the circles). When you are finished, hand in just the answer sheets. You can keep the question sheets. There are 15 questions, each worth 5 points. Everyone receives 25 points for free. Good Luck!

1) WRITTEN
2) WRITTEN
3) WRITTEN
4) WRITTEN
5) WRITTEN
6) (A) (B) (C) (D) (E)
7) (A) (B) (C) (D) (E)
8) (A) (B) (C) (D) (E)
9) (A) (B) (C) (D) (E)
10) (A) (B) (C) (D) (E)
11) (A) (B) (C) (D) (E)
12) (A) (B) (C) (D) (E)
13) (A) (B) (C) (D) (E)
14) (A) (B) (C) (D) (E)
15) (A) (B) (C) (D) (E)
Answer For Question 1:
Answer for Question 2:

Answer for Question 3:
Answer for Question 4:

Answer for Question 5:
1) The following table presents data collected in the 1960s for 21 countries on X=Annual Per Capita Cigarette Consumption (“Cigarette”), and Y=Deaths from Coronary Heart Disease per 100,000 persons of age 35-64 (“Coronary”).

<table>
<thead>
<tr>
<th>Country</th>
<th>Cigarette</th>
<th>Coronary</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>3900</td>
<td>259.9</td>
</tr>
<tr>
<td>Canada</td>
<td>3350</td>
<td>211.6</td>
</tr>
<tr>
<td>Australia</td>
<td>3220</td>
<td>238.1</td>
</tr>
<tr>
<td>New Zealand</td>
<td>3220</td>
<td>211.8</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>2790</td>
<td>194.1</td>
</tr>
<tr>
<td>Switzerland</td>
<td>2780</td>
<td>124.5</td>
</tr>
<tr>
<td>Ireland</td>
<td>2770</td>
<td>187.3</td>
</tr>
<tr>
<td>Iceland</td>
<td>2290</td>
<td>110.5</td>
</tr>
<tr>
<td>Finland</td>
<td>2160</td>
<td>233.1</td>
</tr>
<tr>
<td>West Germany</td>
<td>1890</td>
<td>150.3</td>
</tr>
<tr>
<td>Netherlands</td>
<td>1810</td>
<td>124.7</td>
</tr>
<tr>
<td>Greece</td>
<td>1800</td>
<td>41.2</td>
</tr>
<tr>
<td>Austria</td>
<td>1770</td>
<td>182.1</td>
</tr>
<tr>
<td>Belgium</td>
<td>1700</td>
<td>118.1</td>
</tr>
<tr>
<td>Mexico</td>
<td>1680</td>
<td>31.9</td>
</tr>
<tr>
<td>Italy</td>
<td>1510</td>
<td>114.3</td>
</tr>
<tr>
<td>Denmark</td>
<td>1500</td>
<td>144.9</td>
</tr>
<tr>
<td>France</td>
<td>1410</td>
<td>144.9</td>
</tr>
<tr>
<td>Sweden</td>
<td>1270</td>
<td>126.9</td>
</tr>
<tr>
<td>Spain</td>
<td>1200</td>
<td>43.9</td>
</tr>
<tr>
<td>Norway</td>
<td>1090</td>
<td>136.3</td>
</tr>
</tbody>
</table>
**Regression Analysis**

The regression equation is

\[
\text{Coronary} = 29.5 + 0.0557 \text{ Cigarette}
\]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>StDev</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>29.45</td>
<td>29.48</td>
<td>1.00</td>
<td>0.330</td>
</tr>
<tr>
<td>Cigarette</td>
<td>0.05568</td>
<td>0.01288</td>
<td>4.32</td>
<td>0.000</td>
</tr>
</tbody>
</table>

S = 46.56 \quad R-Sq = 49.6\% \quad R-Sq(adj) = 46.9\%

**Analysis of Variance**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>40484</td>
<td>40484</td>
<td>18.68</td>
<td>0.000</td>
</tr>
<tr>
<td>Residual Error</td>
<td>19</td>
<td>41181</td>
<td>2167</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>20</td>
<td>81666</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A) Based on the scatterplot of Y versus X, does there appear to be a linear relationship between cigarette consumption and heart disease? If so, does the relationship appear to be negative or positive?

B) What patterns or problems, if any, do you see in the residuals versus X plot? Would you feel reasonably comfortable in fitting a simple linear regression model to this data set?

C) Write the equation for the fitted model.

D) Give an interpretation of the fitted slope, \( \hat{\beta}_1 \).

E) How much natural variability is associated with \( \hat{\beta}_0 \)?

2) For the situation described in Problem 1, answer these questions.

A) Based on the Minitab output, is it plausible that the true intercept \( \beta_0 \) is zero? Explain. What would be the practical interpretation of the result that \( \beta_0 = 0 \)? Is there any contradiction here?

B) Do you think that natural variability alone could account for such a large value of \( \hat{\beta}_1 \) as actually found here? Explain.

C) Using the Minitab output, determine whether sufficient statistical evidence exists to conclude that there is a positive linear relationship between X and Y at the 1% level of significance.

D) Based on \( R^2 \), assess the strength of the linear relationship between X and Y.
E) Do the $p$-value for $\hat{\beta}_1$ and the value of $R^2$ provide contradictory evidence on the strength of the linear relationship between smoking and heart disease? Explain.

3) The weights of ten $100$ casino chips (selected at random from a large batch of new $100$ chips at the Golden Nugget Casino) averaged $0.8$ ounces, with a sample standard deviation of $0.03$ ounces.

A) Assuming that the weights of the chips in the batch are normally distributed, construct a $95\%$ confidence interval for the mean weight of the entire batch.
B) Does the interval you got in part A) have a $95\%$ chance of containing the mean weight of the entire batch? Explain.

4) For the situation described in Problem 3, if $\mu$ is the mean weight for the entire batch,

A) Test $H_0 : \mu = 0.83$ versus $H_1 : \mu \neq 0.83$ at level $.05$.
B) What is the probability that $H_0$ is true? Explain.

5) One hundred randomly selected milk cows were observed for one week and then given a genetically engineered drug designed to increase milk production. The increase in milk production (second week minus first week) averaged to $11$ gallons with a sample standard deviation of $50$ gallons.

A) State the appropriate null and alternative hypotheses for this problem, in terms of $\mu$.
B) What is the meaning of $\mu$ (in terms of cows)?
C) What do the null and alternative hypotheses imply about the effectiveness of the drug?
D) Give all values of $\alpha$ at which the null hypothesis can be rejected.
E) Suppose the drug had no effect. Then out of $1000$ random samples of $100$ cows, how many samples would be expected to yield an increase in milk production at least as large as what was found in our sample?
Questions 6-10 concern the following situation. A random sample of 50 adults were asked how much they spend on lottery tickets, and were interviewed about various socioeconomic variables. The variables are

- PercLott = Percentage of total household income spent on the lottery. (This is Y).
- YrsEdu = Number of years of education,
- Age = The person’s Age,
- Kids = Number of Children,
- Income = Personal income (Thousands of Dollars).

Here is the Minitab regression output:

**Regression Analysis**

The regression equation is

\[ \text{PercLott} = 15.1 - 0.591 \text{YrsEdu} + 0.0065 \text{Age} + 0.082 \text{Kids} - 0.0666 \text{Income} \]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>StDev</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>15.070</td>
<td>2.444</td>
<td>6.17</td>
<td>0.000</td>
</tr>
<tr>
<td>YrsEdu</td>
<td>-0.591</td>
<td>0.1813</td>
<td>-3.26</td>
<td>0.002</td>
</tr>
<tr>
<td>Age</td>
<td>0.00647</td>
<td>0.03395</td>
<td>0.19</td>
<td>0.850</td>
</tr>
<tr>
<td>Kids</td>
<td>0.0816</td>
<td>0.2665</td>
<td>0.31</td>
<td>0.761</td>
</tr>
<tr>
<td>Income</td>
<td>-0.06663</td>
<td>0.03305</td>
<td>-2.02</td>
<td>0.050</td>
</tr>
</tbody>
</table>

\[ S = 2.389 \quad \text{R-Sq} = 61.2\% \quad \text{R-Sq(adj)} = 57.7\% \]

**Analysis of Variance**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>4</td>
<td>404.42</td>
<td>101.10</td>
<td>17.72</td>
<td>0.000</td>
</tr>
<tr>
<td>Residual Error</td>
<td>45</td>
<td>256.80</td>
<td>5.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>49</td>
<td>661.22</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6) Based on the output, is there statistical evidence to suggest that relatively educated people spend less on lotteries than relatively uneducated people?  
   A) Yes  B) No

7) Is there statistical evidence to suggest that older people spend more on lotteries than younger people?  
   A) Yes  B) No

8) The results of the F test imply that, beyond a reasonable doubt:  
   A) All of the true slope coefficients in the model are nonzero  
   B) At least one of the true slope coefficients in the model is nonzero  
   C) None of the true slope coefficients in the model is nonzero  
   D) All of the estimated slope coefficients are nonzero  
   E) At least one of the estimated slope coefficients is nonzero
9) The 95% confidence interval for the true coefficient of YrsEdu is
A) (–2.12, 3.14)  B) (–0.5911, 0.5911)  C) (–1.1)  D) (–0.946, –0.236)
E) (–1.06, –0.124).

10) Performing a two-tailed hypothesis test for the null hypothesis that the true
coefficient of YrsEdu is –1, at the 5% level of significance, we:
A) Reject the null hypothesis
B) Do not reject the null hypothesis

11) Let’s return to the simple regression described in Problem 1. The residual for
Greece is:
A) 1800  B) 29.45  C) 31.74  D) 1768.26  E) –88.474

12) In linear regression, does a point with high leverage necessarily cause the
fitted line to change?
A) Yes  B) No

13) In an election last year, a politician received 58% of the ballots cast. Several
months later, a survey of 700 people revealed that 54% now support her. Is
this sufficient evidence to allow us to conclude at the .05 level of significance,
that her popularity has decreased?
A) Yes  B) No

14) If the probability of rain tomorrow is 1/3, then the odds that it will rain
tomorrow are:
A) 1:1  B) 3:1  C) 2:1  D) 3:2  E) 6:5

15) A sample of size 100 is going to be taken from a population with mean 3 and
variance 25. The probability that the sample mean will exceed 4 is:
A) .0456  B) .4207  C) .0793  D) .4772  E) .0228