1. Biologists anesthetized 14 wild bears and took various physical measurements on each, including neck circumference in inches and chest circumference in inches. Below is SAS output from an analysis of these data.

(a) (i) We wish to try to predict chest circumference when we know neck circumference. Which variable is the response variable?

(b) (i) On the scatterplot, circle the point(s), if any, that appear to be outliers.

(c) (i) What is the direction of the relationship (positive/negative)?

Plot of chest vs. neck. Symbol used is ‘.’.

1. Show your work on any problem that involves calculations.

Name: ________________________________ Course no. (30 or 105) __________

1. Biologists anesthetized 14 wild bears and took various physical measurements on each, including neck circumference in inches and chest circumference in inches. Below is SAS output from an analysis of these data.

(a) (i) We wish to try to predict chest circumference when we know neck circumference. Which variable is the response variable?

(b) (i) On the scatterplot, circle the point(s), if any, that appear to be outliers.

(c) (i) What is the direction of the relationship (positive/negative)?

Plot of chest vs. neck. Symbol used is ‘.’.

---

The REG Procedure
Model: MODEL1
Dependent Variable: chest

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>4050.46867</td>
<td>4050.46867</td>
<td>360.27</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>42</td>
<td>484.63726</td>
<td>11.243D2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>53</td>
<td>4535.10593</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Root MSE 3.35306 R-Square 0.9739
Dependent Mean 35.66296 Adj R-Sq 0.9714
Coeff Var 9.40508

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > |t |
|----------|----|--------------------|----------------|---------|-------|
| Intercept| 1  | 3.80568            | 1.79333        | 2.19    | 0.032 |
| neck     | 1  | 1.54982            | 0.08165        | 18.98   | <.0001|

Parameter Estimates

<table>
<thead>
<tr>
<th>Variable</th>
<th>DF</th>
<th>95% Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>0.31535</td>
</tr>
<tr>
<td>neck</td>
<td>1</td>
<td>1.38597</td>
</tr>
</tbody>
</table>

(d) (iii) If bear A’s neck is 2 inches larger than bear B’s neck, how much larger would you expect bear A’s chest to be than bear B’s chest?
(c) State the null hypothesis of no linear relationship between neck circumference and chest circumference in terms of the appropriate regression parameter or parameters. Define any symbols that you use.

(f) Write the corresponding two-sided alternative hypothesis.

(g) Can we reject the null hypothesis at the $\alpha = 0.01$ significance level? Tell how you used the SAS output to justify your answer.

(b) SAS gives a 95% confidence interval of (1.36397, 1.71367) for the coefficient of neck. This means that we are 95% confident that (circle one):

i. the sample mean chest circumference is in this interval

ii. the sample slope capturing the relationship between chest and neck circumference is in this interval

iii. the population mean chest circumference is in this interval

iv. the population slope capturing the relationship between chest and neck circumference is in this interval

v. the expected chest circumference of a bear with neck circumference of 0 inches is in this interval

vi. we don't want to measure bears' chest circumferences ourselves

vii. none of the above

2. The Washington Post reported on 4/25/03 on numbers released by the World Health Organization the previous day regarding the numbers of confirmed cases of SARS and the numbers of deaths from SARS worldwide and in specific countries. The death rate may be defined as the proportion of all people infected with a disease who will die from that disease. I wish to use the data from the article to consider whether the death rate from SARS is different in Canada from in the rest of the world. The SAS output below summarizes the numbers in the Washington Post article.

```
(a) (1) We are interested in comparing the death rate in the population of all Canadians infected with SARS to the death rate in the population of all people infected with SARS in the rest of the world. First, what is the point estimate for the population proportion of infected Canadians who will die of SARS? (numeric answer)

(b) (3) Are the rules of thumb met so that we can use the normal approximation to construct a 95% confidence interval for the population proportion of infected Canadians who will die of SARS? Briefly state the rules of them and say whether each is met with this dataset.
```
(c) Construct a 95% confidence interval for the population proportion of infected Canadians who will die of SARS. (numeric answer)

(d) Under the null hypothesis that the death rate is the same in Canada as in the rest of the world, what is the point estimate of the common proportion worldwide who will die?

(e) Construct the table of expected counts under the null hypothesis that the death rate is the same in Canada as in the rest of the world, using conventional statistical symbols.

<table>
<thead>
<tr>
<th>Region</th>
<th>Status</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada</td>
<td></td>
<td>140</td>
<td>3.18</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td>4299</td>
<td>96.82</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>4439</td>
<td>100.00</td>
</tr>
</tbody>
</table>

(f) Below is SAS output for the chi-square test for this table. Write the formal expression for the null and alternative hypotheses using conventional statistical symbols.

(g) Can you reject the null hypothesis at the .05 significance level? Why or why not?

<table>
<thead>
<tr>
<th>Statistic</th>
<th>DF</th>
<th>Value</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Square</td>
<td>1</td>
<td>5.9496</td>
<td>0.0147</td>
</tr>
</tbody>
</table>

3. We have obtained the dataset described below.

DESCRIPTIVE ABSTRACT:
For 97 countries in the world, data are given for birth rates, death rates, infant death rates, life expectancy at birth for males, life expectancy at birth for females, and Gross National Product.

VARIABLE DESCRIPTIONS:
Columns
1 - 6 Live birth rate per 1,000 of population
7 - 14 Death rate per 1,000 of population
15 - 22 Infant deaths per 1,000 of population under 1 year old
23 - 30 Life expectancy at birth for males
31 - 38 Life expectancy at birth for females
39 - 46 Gross National Product per capita in U.S. dollars
47 - 52 Country Group
   1 = Eastern Europe
   2 = South America and Mexico
   3 = Western Europe, North America, Japan, Australia, New Zealand
   4 = Middle East
   5 = Asia
   6 = Africa
53 - 74 Country

We wish to use these data to determine whether the population means of life expectancy for males differ in these geographical regions identified by the country group variable.

(a) (2) Of the statistical methods listed below, the one most likely to be useful in addressing this question is: (Circle one)
   i. one-sample t-test
   ii. paired sample t-test
   iii. two-sample t-test
   iv. two-sample z-test for proportions
   v. Chi-square test
   vi. ANOVA

(b) (2,5) Consider the assumptions of the method you chose in the previous question. Why might one or more of these assumptions not be met for these data? Explain briefly. (More than one possible right answer here)