CS130 Practice Problems for the Final

Skills you will need:

- Connect to grace
  - Import data from the Web in Excel
  - Import Excel data into R
  - Edit/cleanup data in Excel to prepare it for R
  - Perform statistical tests in R
  - Write R scripts
  - Use Excel and R to visualize data in a meaningful way
  - Paste results into a Word document as specified (similar to assignment #3)

Work through as many of the following problems before 6pm tomorrow as possible and save the results in a Word document called PUNetIDReview. Some problems specify Excel or R. If neither is specified you may choose or you could do it in both!

Remember, the key to R is determining how to layout the data and which test to perform.

Problem #1:

For a random number of selected homes sold in Washington County, the annual tax amount (in thousands of dollars) and selling price (in thousands of dollars) are as follows.

Taxes:  4.0  2.4  1.8  1.5  1.4  1.4  3.0  1.9

SellingPrice:  265  142  114  160  130  150  228  145

What constitutes a single case or item? Remember, this will be represented by a row.

___________________

What are the variables? Remember, this will be represented by columns.______________________________

Perform a Linear Regression on the above data using Excel and paste your results to a Word document. Make sure the linear equation and R-squared value is displayed.

What is the independent variable? _________________What is the dependent variable? ______________

Does the regression equation represent a strong correlation? Why?

For an annual tax amount of $3,400, what might we expect the selling price of this house to be?

For a selling price of $335,000, what might we expect the taxes to be for this house?
Problem #2:
The worksheet Acetaminophen.xlsx in CS130Public on grace contains data for two cold medicines tested for acetaminophen.

Test the claim that the mean amount of acetaminophen is the same for both medicines. You are to create an R script acetaminophen.R that loads the Excel data from Acetaminophen.xlsx into a data frame and outputs the results of the proper test when run. Assume the Excel file will be read from your Desktop.

What constitutes a single case or item? ____________________
What are the variables? ______________________________

Explain what Hypothesis test you will be using to solve the problem.

State the Null Hypothesis.

Perform the proper Hypothesis test with a 95% confidence level.

State your conclusion and reasoning for your conclusion.

Problem #3:
The NCAA is considering ways of speeding up the end of college basketball games. In a recent NCAA tournament, the last two minutes of 60 games took the following times to complete (in seconds):

756  587  929  871  378  503  564  1128  693  748
448  670  1023  335  540  853  852  495  666  474
443  325  514  404  820  915  793  778  627  483
861  337  292  1070  625  457  676  494  420  862
991  615  609  723  794  447  704  396  235  552
626  688  506  700  240  363  860  670  396  345

Note: Data was reported in USA Today

a) In both Excel and R, find the mean, median, standard deviation, minimum, and maximum values for the given data.

b) Create a histogram of the given data in R.

c) Do these times indicate that something should be done to speed up the last two minutes of the game (is the last two minutes too long)? Support your answer as best you can.
**Problem #4:**
Radioactive elements decay into non-radioactive elements at a very slow rate. We talk about the half-life of an element which is the amount of time it takes half of the substance to decay into a non-radioactive element.

We can use the formula \( A(t) = A(0) \times (1/2)^{(t/h)} \) to determine how much of a radioactive substance we have left after time \( t \). \( A(0) \) is the initial amount of radioactive material. \( h \) is the half-life of the element.

1) Create a worksheet that allows the user to input an initial amount and half-life and displays a table showing the time \( t \) and the amount of the element left after time \( t \) from time 0 to time 100 by increments of 5 years. Use named cells in your worksheet. The amount of initial substance is 100 grams and the half-life is 25 years.

2) Create a scatterplot (with proper labeling) of the amount of radioactive substance left over the course of time. Include a regression equation and an \( R^2 \) value. Use exponential for the trendline.

3) At what time can we expect 50% of the radioactive material to remain? To perform your calculation, first solve the exponential equation in terms of \( x \). Then do the calculation. Check your answer with Goal Seek.

**Problem #5:**
The NCAA is considering ways of speeding up the end of college basketball games. In a recent NCAA tournament, the last two minutes of 60 games took the following times to complete (in seconds):

<table>
<thead>
<tr>
<th>Time (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>756 587 929 871 378 503 564 1128 693 748</td>
</tr>
<tr>
<td>448 670 1023 335 540 853 852 495 666 474</td>
</tr>
<tr>
<td>443 325 514 404 820 915 793 778 627 483</td>
</tr>
<tr>
<td>861 337 292 1070 625 457 676 494 420 862</td>
</tr>
<tr>
<td>991 615 609 723 794 447 704 396 235 552</td>
</tr>
<tr>
<td>626 688 506 700 240 363 860 670 396 345</td>
</tr>
</tbody>
</table>

Note: Data was reported in USA Today

a) In both Excel and R, find the mean, median, standard deviation, minimum, and maximum values for the given data. Paste these results into a Word document.

b) Create a histogram of the given data in R.

c) Do these times indicate that something should be done to speed up the last two minutes of the game (is the last two minutes too long)? Support your answer as best you can.
**Problem #6:**

A car company CarsRUs claims their sedans average 31 mpg. You randomly select 10 sedans from local dealerships and test their gas mileage under similar conditions. Here are the mpg scores: 30, 28, 25, 33, 26, 26, 28, 30, 33, 31

State the Null Hypothesis.

Perform the proper Hypothesis test with 95% confidence level and paste in the results.

State your conclusion and reasoning for your conclusion.

**Problem #7:**

You have been hired at a water treatment plant to manage the water quality of the treated water. A new set of filters are installed and you want to determine if they were worth the money. You measure the bacteria count in treated water before and after the new filters are installed. The data you collected is shown below, two samples for each of the 11 clean water tanks. Did the new filter reduce the amount of bacteria in the clean water tanks?

<table>
<thead>
<tr>
<th>Tank #</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old Filter</td>
<td>10</td>
<td>15</td>
<td>1</td>
<td>3</td>
<td>10</td>
<td>2</td>
<td>18</td>
<td>16</td>
<td>8</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>New Filter</td>
<td>4</td>
<td>10</td>
<td>2</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

State the Null Hypothesis.

Perform the proper Hypothesis test with 95% confidence level and paste in the results.

State your conclusion and reasoning for your conclusion.

**Problem #8:**

You have been offered a 5 year loan on a $10,000 car with an 8.7% annual interest rate compounded monthly. What will your monthly payment be?

You can afford a $300 monthly payment. How does that affect the price of the car you can purchase using this same 5 year loan with an 8.7% interest rate?

**Problem #9:**

Your very good (but sneaky) friend offers to loan you $5,000 for a new car. Your friend is willing to give you the loan for 5 years at 6.5% annual interest or 3 years at 8% annual interest. Which option would you choose? Support your decision with data!