**Scenario:**

To succeed in any industry that deals with massive amount of data, you must possess analytic skills. As you interview for an entry level position after receiving your undergraduate/graduate degree, let us say, you are given a series of four different analyses to ***conduct*** and ***interpret***.

**Objective:**

Complete the tasks and corresponding questions given below. *All of the datasets are on 4 different sheets (named as PROBLEM1, PROBLEM2, PROBLEM3 & PROBLEM4, respectively) within ONE Excel worksheet,* titled ***Project3.xlsx***.

The data will need to be imported into SAS prior to analysis. Some data have not been given to you in a format conducive to analysis and may need to be modified in order to conduct the appropriate such analysis.

**Submission Procedure:**

Please submit the following:

1. SAS program or R Codes
2. Word document with your answers

**Problem 1:**

* The dataset contains 2 variables **GROUP** (two levels: **BP\_US** and **BP\_FOREIGN**) and **BLOODPRESSURE**, denoting blood pressures of individuals born within (**BP\_US**) and outside (**BP\_FOREIGN**) of the United States.
* You wish to determine whether or not there is a significant difference in mean blood pressures between the two groups.
* Conduct an independent samples t-test to make this determination. **The data file for problem 1 has 10,000 observations (5000 observations for each level of GROUP)**.

**Problem 2:**

* A study is conducted on mice with a glioblastoma (brain tumor) to determine if a certain drug improves survival.
* The dataset has two variables: **TREATMENT** and **DAYS\_TO\_DEATH**.

*TREATMENT* has 3 groups: CONTROL, TREAT\_A (receives 100 mg of drug), andTREAT\_B (receives 200mg of the drug).

* Use a one-way ANOVA to determine if there is a significant difference in survival among the three groups.
* If significant differences exist, use the Bonferroni test to determine specifically which pair of treatment groups differs. **The data file for problem 2 has 300 observations (each group has 100 observations)**.

**Problem 3:**

* Assuming a causal relationship, can CALORIES intake be used to predict systolic blood pressure (**SYSTOLIC\_BP**)? **The data file for problem 3 has 200 observations**.
* Use SAS to perform a simple linear regression analysis and obtain the regression line. Based on the PROC REG output,

**Problem 4:**

* There are two variables in the final dataset. The variable, **BTHDEFECT** equals 1 if a child was born with a birth defect and 0 if not.
* Likewise, the variable, **PRETERM,** equals 1 if a child was born premature and 0 if not.
* Perform a chi-square analysis to determine if there is an association between *preterm birth* and *having a birth defect*. **The data file for problem 4 has 10,000 observations**.

**Answer the following questions and submit the Word document. Please type in your answer into this Word document.**

**PROBLEM 1:**

1. **Is there a statistically significant difference in the mean blood pressure between the two groups? Give reason for your answer.**

**PROBLEM 2:**

1. **Does the variable, DAYS\_TO\_DEATH, have normal distribution for each level of the TREATMENT? Give reasons for your answer.**
2. **Is there a statistically significant differences in the mean *days\_to\_death* among the three groups? Give reason for your answer.**
3. **Which pair of the treatment groups differ statistically significantly? Use Bonferroni test.**

**PROBLEM 3:**

1. **Obtain the linear regression equation.**
2. **What percent of the variability observed in the systolic blood pressure scores can be accounted for by a linear relationship between systolic blood pressure and calorie intake?**
3. **Fill in the blank. As predicted by the model, for a 100-unit increase in calorie consumption, the mean systolic blood pressure \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.** 
   * + 1. **(HINT: think about the change in mean systolic blood pressure for a 1-unit increase and multiply by 100)**
4. **Would you reject or fail to reject the null hypothesis that calorie intake is associated with systolic blood pressure? Why or why not?**

**PROBLEM 4:**

**1. What is the interpretation of the results based on the SAS output?**