**MATH29264 Data Modelling Project**

*Overview*

Consider the data compiled by the Canadian Cancer Society on Cancer Statistics in 2021, use this data to compile, organize and analyze the trends on incidence rate, mortality rate and survival. This document outlines the steps required to complete this project. The summary figures and raw data tables associated with this assignment are available on SLATE. You will need to use Excel or another statistical modelling software to complete this assignment.  
  
 This project is worth **15% of your Grade and is due no later than Saturday June 18TH @ 11:59 PM**

*Section 1 – Incidence Rates*

1. From the Incidence by Sex data, classify the data in Figure 1.2 as qualitative or quantitative and identify the level of measurement (Classify both variables) **(2 marks)**
2. Select at least 7 non-sex specific cancers and using an appropriate graph, display the data for both sexes on the same figure. Select two types of cancers from your graph and describe the differences in the data. **(3 marks)**

*Section 2 – Mortality Rates*

1. From Figure 2.3, assuming a normal distribution, calculate the probability that a cancer related death occurs below the age of 65. **(2 marks)**
2. Calculate the age range that accounts for the middle 50% of deaths. **(2 marks)**
3. Using the incidence rate by age data for a single gender, plot this data as a scatter plot, identifying the independent and dependent variables. **(2 marks)**
4. Fit the data with an appropriate mathematical model (linear or non-linear) and calculate the equation of the line. **(2 marks)**
5. Create a log-linear model of the data and graph this data using a scatter plot. Calculate the equation of the line and the coefficient of determination and correlation coefficient.  
   **(4 marks)**

*Section 3 – Survival Rates*

1. For the prostate or breast cancer data in Figure 3.1, fit a linear model to the data and calculate the equation of the line, correlation coefficient and coefficient of determination for this data. **(3 marks)**
2. Using the equation of the line, estimate the 15yr survival rate for the cancer you chose.  
   **(1 mark)**
3. Why has there been an improvement in the survival for breast and prostate cancer? Use an appropriate reference to support your answer. **(2 marks)**

*Additional Criteria*

1. Appropriate citation style and reference list **(1 mark)**
2. Neatness, presentation of graphs and data, all graphs have axis labels and appropriate titles. **(1 mark**)