1. For the following research questions, identify the correct regression you would need to use (and subtype of linear regression, if applicable) and identify the independent and dependent variables, as well as whether they are continuous or categorical values: (4 marks)

a) Joseph would like to know what demographic factors (i.e., sex, age, race) are related to colorectal tumor diameter in cancer patients.

This situation needs to use statistical multiple regression. Colorectal tumor diameter (continuous value) is the dependent variable, and sex (categorical values), age (continuous or categorical values), race (categorical values) are the independent variables.

b) Priya is interested in determining whether the number of drivers pulled over for speeding is impacted by the day of the week in Ontario.

This situation needs to use standard multiple regression to test a theory. The day of the week in Ontario (categorical value) is independent variable, and the number of drivers pulled over (continuous value) for speeding is dependent variable.

c) Josie would like to investigate whether consideration of sleep duration improves the prediction of academic grades above and beyond the duration of studying hours.

This situation needs to use hierarchical multiple regression to test a theory. Academic grades (continuous value) is dependent variable, while sleep duration (continuous value) and the duration of studying hours (continuous value) are independent variables.

d) Kevin would like to be able to predict whether older adults will snore or not based on their age group (50s, 60s or 70s), gender, and type of pillow (soft or hard).

This situation needs to use logistic regression. Older adults will snore or not is dependent variable, and their age group (50s, 60s or 70s) (categorical values), gender (categorical values), and type of pillow (soft or hard) (categorical values) are independent variables.

2. The following is output from a Poisson regression that investigated whether the number of cases of esophageal cancer is related to the age group (20s, 30s, 40s, 50s, 60s, 70s) and alcohol consumption (Low, Occasional, Frequent, Excessive) of an individual. (5 marks)

Identify the following pieces of information:

a) What does the estimate of the Intercept indicate?

This is the Poisson regression estimate when all variables in the model are evaluated at zero. It represents the number of counts with the base predictors.

b) What do significant ($p<0.05$) p-values of the coefficients indicate?

The significant ($p<0.05$) p-values of the coefficients indicate that the correlated predictors are significant for the independent.

c) In a written description, interpret the output for age_A60 coefficient.
A 1 unit increase in age_A60 is associated with a exp(age_A60 coefficient)=exp(1.85)=6.36 times more in the number of cases of esophageal cancer.

d) I ran a second Poisson regression without considering alcohol consumption and found the AIC value to be 384.33. What does this tell you about the relative strengths of the models?

When compare to models with the same dependent outcome, AIC value lower means the model is better. However, in this case, only one model gives the AIC value, so we cannot tell which one is better.