Project 5: Make It Make Sense!

Purpose

The project aims to develop a model predicting survival time after kidney surgery, exploring the relationship between kidney function and survival time, and examining if this relationship is affected by alcohol consumption

Data

The data comprises patient records from a random sample of individuals who underwent the same kidney surgery. This analysis focuses on survival time after surgery until death of a patient (median = 656.5 days, IQR = 454.25 days), kidney function score (median = 2.4, IQR = 1.875), and alcohol use (0=none, 1=moderate, 2=severe).

Methods

Below is the full regression model used to estimate survival time based on kidney function score, alcohol use, and the interaction between kidney function score and alcohol use. The log transformation was done to meet all four assumptions necessary when applying a multiple linear regression model.

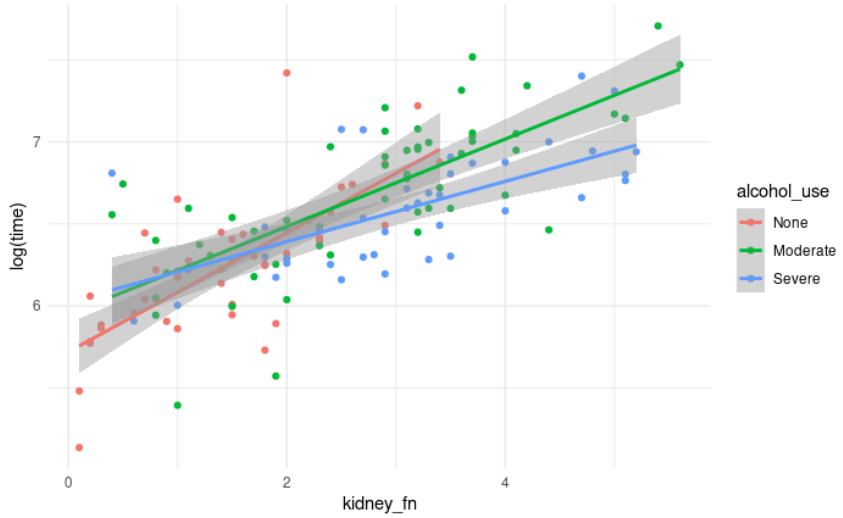
$$log(\text{time})\text{-hat} = \beta 0 + \beta 1 * \text{kidney\_fn} + \beta 2 * \text{alcohol} + \beta 3 * (\text{kidney\_fn} * \text{alcohol\_none}) + \varepsilon$$

A hypothesis test was conducted to assess whether kidney function score interacts with alcohol use.

$$H0: \beta \text{kidney\_fn} * \text{alcohol} = 0$$
$$Ha: \beta \text{kidney\_fn} * \text{alcohol} \neq 0$$

According to the t-test, with a p-value of 0.00553 (t=-2.820, df=134), there is evidence of an interaction between kidney function score and alcohol use.

Figure 1: Interaction graph between kidney function score and alcohol use



Since the p-value for the full model is 0.5909 (F = 66.95, df = 3 and 134) while the t-test for alcohol is high (t=2.578, p-value 0.011018). The variable alcohol is removed from the model leaving:

$$log(\text{time})\text{-hat} = \beta 0 + \beta 1 * \text{kidney\_fn} + \beta 2 * (\text{kidney\_fn} * \text{alcohol\_use}) + \varepsilon$$

Results

The final estimated regression model to predict time, according to kidney function score, alcohol use, and interaction between them is as follows based off of the regression out:

$$log(\text{time})\text{-hat} = 5.873 + 0.256 * \text{kidney\_fn} + 0.107 * \text{alcohol\_moderate} - 0.069 * \text{alcohol\_severe}$$

Conducting an F-test, with a p-value of 0.59 (F=40.13, df = 5 and 132), this model is useful in predicting survival time.

As a result, For each 1 point increase in kidney function, the log-transformed survival time increases by 0.36294. For patients with moderate alcohol use (compared to no alcohol use), the expected log-transformed survival time increases by 0.23031. For patients with severe alcohol use (compared to no alcohol use), the expected log-transformed survival time increases by 0.30356. For each 1 point increase in kidney function, the survival time is reduced by 0.096 days for those with moderate alcohol consumption. The effect of kidney function on survival time is reduced by 1.79 days for patients with severe alcohol use.