

Project 5 - Minh and Oliver

To predict the estimated survival time for patients after kidney surgeries, we can consider using a patients' kidney function score and the interaction it has with zero, moderate, or severe alcohol use. This interaction between kidney score and alcohol effects can predict the number of days from the patients' surgery until death.

The chosen model to predict estimated survival time used variables: kidney function, moderate alcohol use (dummy), severe alcohol use (dummy) and the two interaction terms for kidney function and alcohol consumption level. Our null hypothesis is that no interaction terms are useful in predicting estimated survival. Our alternative hypothesis is that at least one interaction term is useful in predicting estimated survival time. A partial F-test was used to test whether it is necessary to include the interaction between kidney function and alcohol use.

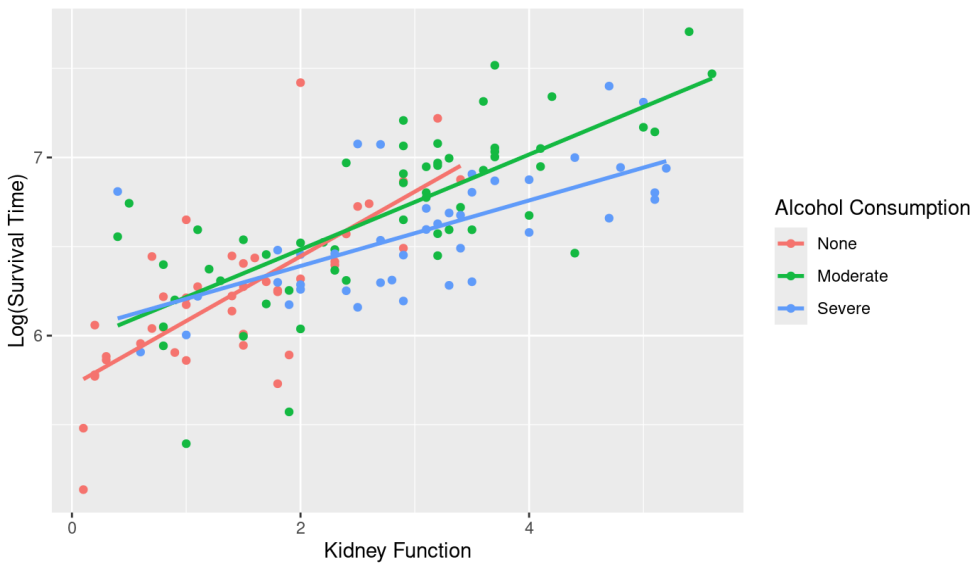
A full model with all main variables and interaction terms was tested against a reduced model with just the main variables. With a p-value of 0.021, less than 0.05, ($F=3.97$, $df=2$ and 134), there is strong evidence that including the interaction between alcohol use and kidney function is useful in determining survival time. Assumptions were checked for the full model with no concerning violations. Table 1 shows the full parameters of the model.

Table 1 - Parameter Estimates for Survival Time Model

Variable	Estimate	SE	t-statistic	p-value
Intercept	5.71936	0.08782	65.126	< 0.0001
Kidney Function	0.36294	0.05226	6.945	<0.0001
Moderate Alcohol	0.23031	0.12974	1.775	0.07818
Severe Alcohol	0.30356	0.15204	1.996	0.04794
Kidney Function * Moderate Alcohol	-0.09620	0.06099	-1.577	0.11715
Kidney Function * Severe Alcohol	-0.17895	0.06449	-2.775	0.00633

There is a clear relationship between kidney function and survival time. When alcohol consumption levels are factored in, slope and intercepts vary from each consumption level. Figure 1 shows this relationship:

Figure 1 - Kidney Function vs Log(Survival Time) by Alcohol Consumption



Based on the scatterplot, the estimated survival time of patients with zero alcohol assumptions has a lower baseline, but exceeds patients with moderate and severe alcohol use as kidney function score increases.

Due to the three reported levels of alcohol, three models can be constructed for each individual case. Below are the three possible regression lines based off of their respective alcohol levels.

No alcohol use: **$\log(\text{Survival Time}) = 5.72 + 0.36(\text{Kidney Function})$**

Moderate alcohol use: **$\log(\text{Survival Time}) = 5.95 + 0.26(\text{Kidney Function})$**

Severe alcohol use: **$\log(\text{Survival Time}) = 6.02 + 0.18(\text{Kidney Function})$**

For someone with no alcohol consumption, every one score increase in kidney function increases the estimated survival time by 43.3%. For those with moderate alcohol usage, a one score increase of kidney function increases the estimated survival time by 29.7%. Finally, for those with severe alcohol usage, for every one additional score of kidney function, the estimated survival time increases by 19.7%. As seen with all slopes for Kidney function, the effect on estimated survival time is always positive regardless of the level of alcohol usage. However, as alcohol usage increases, the Kidney function's effect on estimated survival time is reduced. Kidney function's effect is largest when no alcohol is consumed, and is the smallest with severe alcohol use.