**SDM Assignment 6 – Sai Kiran Batchu**

The given data is a survival model type with survival days and status as the y variables.

***Survival in Days Dependent variable Predictor Table***

|  |  |  |
| --- | --- | --- |
| **Predictor** | **Effect** | **Rationale** |
| Treatment | +/- | New treatment types can have either positive or negative effects on the patient’s survival |
| Cell type | +/- | Squamous, adeno, and large cell types often have more survival rates than small cell typed cancers |
| Karnofsky score | + | More Karnofsky score can indicate a higher survival rate as the person is healthier |
| Months from Diagnosis | - | More months of diagnosis can yield lower survival rates as cancer must have grown by that time unless treated properly |
| Age in years | - | An aged person can have a lower survival chance of cancer |
| Prior chemotherapy | +/- | It can have a positive or negative impact on the person’s survival as the cancer cells can be milder / harsher if diagnosed again. |

***Feature Engineering***

Cell\_type: from my study, I found that small cell cancer type is more deadly as the spread is very faster and there are chances of getting cancer again. Mostly squamos, adeno and large cell types are considered as non-small cell type lung cancers so I grouped them into non\_small\_cell type.

***Models***

I have created 2 parametric, 1 semi-parametric, and a non-parametric model to study the marginal effects of relevant predictors and survival graphs.

y = Surv(survival\_days,status)

**Non Parametric:** km2 = survfit(y~ treatment\_type)

**Semi-Parametric:** cox = coxph(y ~ treatment\_type + age + cell\_type + karnofsky\_score + prior\_chemo + months\_from\_diagnosis)

**Parametric:**exp = survreg(y ~ treatment\_type + age + cell\_type + karnofsky\_score + prior\_chemo + months\_from\_diagnosis, dist="exponential")

**Parametric:** weibull = survreg(y ~ treatment\_type + age + cell\_type + karnofsky\_score + prior\_chemo + months\_from\_diagnosis, dist="weibull")

***Summary of non-parametric models***

**Treatment type 0 = standard treatment**

**Treatment type 1 = test treatment**

Chart, histogram

Description automatically generated***A picture containing table

Description automatically generatedA picture containing table

Description automatically generated***

**Kaplan Meier Survival Graph**

***Stargazer summary of the Parametric and semi-parametric models***

***Text

Description automatically generated***

***Graphical user interface

Description automatically generated***

*1. We would like to see Kaplan-Meier survival graphs for patients with the test vs standard treatment. Use this data to assess:*

1. *What is the probability that the patient will survive for 6 months (183 days) and 1 year (365 days) on the standard treatment vs the test treatment?*

**Treatment group Standard**: the probability of a patient surviving for 6 months is approximately 21.24% (taking 177 days from the KM output) and the probability of a patient surviving for 365 days is approximately 53.1% (taking 384 days from the KM output)

**Treatment group Test**: the probability of a patient surviving for 6 months is approximately 21.62% (taking 186 days from the KM output) and the probability of a patient surviving for 365 days is approximately 10.98% (taking 357 days from the KM output)

1. *What is the median number of days where a patient can be expected to survive if they are on the standard vs the test treatment?*

**Treatment group Standard**: The median number of days that a patient will survive is 100 days (probability of 50.20% from the KM output)

**Treatment group Test**: The median number of days that a patient will survive is 50 days (probability of 50% from the KM output)

*2. Create three semi-parametric and parametric models to estimate the marginal effects of relevant predictors on survival outcomes. Interpret the precise effects of standard vs test treatment and other model predictors.*

Interpretation for exponential and Weibull models is the same as the coefficient values are similar.

|  |  |  |
| --- | --- | --- |
| **Predictor** | **CoxPH** | **Exponential and Weibull** |
| Treatment\_type | Test treatment type increases the chance of a person dying by 35.4% | Test treatment type decreases the survival rate by 24.8% |
| Age | If age increases by 1 year the chance of death decreases by 0.7% | Each additional year in age increases the survival rate by 0.3% |
| Cell\_type small\_cell | Small cell type increases the chance of death by 48% | Small cell type decreases the survival rate by 40.8% |
| Karnofsky\_score | If the score increases by 1 number chance of death decrease by 3.3% | If the karnofsky score increases by 1 then the survival rate would increase by 3.3% |
| Prior chemo Yes | Treatment with prior chemo decreases the death rate by 4.5% | Prior chemotherapy treatment would increase the survival rate by 8.3% |
| Months from Diagnosis | If the month from diagnosis increases by 1 then the chance of death decreases by 0.2% | If the month from diagnosis increases by 1 then the chance of survival decreases by 0.01% |