**ISM 6137 - Statistical Data Mining**

**Assignment 6**

**Sahil Shah – 19895141**

**Purpose & Method**

In order to assess the effect of the new treatment, the following analysis was conducted on the data provided. Survival models were built to analyze the treatment’s effect on longevity and the effect of other variables on chance of death.

The Predictor Table below indicates which of the variables were considered the predicted effect.

|  |  |  |
| --- | --- | --- |
| **Variable** | **Effect** | **Reasoning** |
| # Treatment | +/- | Depending on treatment type, life expectancy will change |
| # Cell type | +/- | Depending on type of disease, life expectancy will change |
| # Karnofsky score | + | higher the general performance, higher the life expectancy |
| # Months from Diagnosis | - | the greater time since diagnosis, the lower the life expectancy would be |
| # Age in years | - | the greater the age, the lower the life expectancy would be |
| # Prior chemotherapy | +/- | may increase life expectancy, or may lower it if it degraded overall health sigificantly |
| **Survival in days and Status are not included as the dependent variable y is a combination of the two** | | |

**1. Kaplan-Meier survival graph for patients with the test vs standard treatment & corresponding analysis.**

**Chart, histogram

Description automatically generated**

Figure - Kaplan Meier Curve for Standard (1) and test (2) Treatments

* 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 Type** | **6 Month Survival Probability** | **1 Year Survival Probability** |
| Standard | 21.24% | 7.08% |
| Test | 23.3% | 10.9% |

* 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 Type** | **Median Expected Days of Survival** |
| Standard | 32 – 34 |
| Test | 36 |

**2. Marginal Effects of the Variables Considered on HR (Hazard Ratio) of event (Death) occurrence**

For this analysis, we ran and analyzed multiple parametric and semi parametric models. The interpretations below are based on the output of the semi-parametric coxph model.

|  |  |
| --- | --- |
| **Variable Considered** | **Effect on Event** |
| Treatment 2 | HR of treatment 2 is non-significant as the CI includes 1 |
| Age in Years | HR of Age is non-significant the CI includes 1 |
| Karnofsky Score | Every unit increase in K-Score produces a ~3.23% higher chance of survival |
| Prior Chemo | HR of Prior Chemo is non-significant the CI includes 1 |
| Cell Type 2 | Compared to CT1, CT2 has a 136% higher chance of death |
| Cell Type 3 | Compared to CT1, CT3 has a 230% higher chance of death |
| Cell Type 4 | HR of CT4 is non-significant the CI includes 1 |
| Months From Diagnosis | HR of Months from Diagnosis is non-significant the CI includes 1 |

For those interested in the technical basis for the above interpretation, the appendix includes the output of three different models built for this analysis. This output does not indicate the precise effect of variables, but rather the sign of effect (+/-) on chance of death/time to death.

Further details on precise effects in the parametric models can be provided upon request.

APPENDIX

Table

Description automatically generated