**SDM Lung Cancer**

Analysis of Lung Cancer patients, probability of survival. This analysis is with two groups with below conditions. Below is the overview of the data.

|  |  |  |
| --- | --- | --- |
|  | status | |
| treatment | Alive during observation period | Dead |
| Standard chemotherapy treatment only | 5 | 64 |
| Chemotherapy combined with test drug | 4 | 64 |

Variables from the dataset and their effect on probability of survival.

|  |  |  |
| --- | --- | --- |
| **Variable** | Effect | Rationale |
| Cell Type | ? | **Squamous:** Cancer starts in squamous cells, which are flat cells that line the inside of the airways in the lungs. They are often linked to a history of smoking and tend to be found in the central part of the lungs, near a main airway.  **Small cells:** Cancer grows & spreads more quickly, it responds better chemotherapy. For most people, this cancer will return at some point  **Adeno (Adenocarcinomas):** Cancer in the cells that would normally secrete substances such as mucus. This type of lung cancer occurs mainly in people who currently smoke or formerly smoked. It is more likely to occur in younger people than other types of lung cancer.  **Large Cell:** Cancer can appear in any part of the lung. It tends to grow and spread quickly, which can make it harder to treat.  Each Cancer is different and treatment approaches for these types are very different. We need to see which type has higher survival rate. |
| Months from Diagnosis | + | Earlier they are diagnosed from cancer from the observation period, meaning they are suffering from cancer for more time, this probability of death is more. |
| Age | + | Cancer treatment procedures are very difficult, people with lesser age have more energy to sustain the procedures. probability of death is more |
| Prior chemotherapy | + | The patient is previously already exposed to radiation, means he is already weak and less immunity. |
| Excluded Variables | | |
| Karnofsky score | 0 | This score is like survival probability, thus highly correlated. |

**Kaplan-Meier survival graphs for patients with the test vs standard treatment.**  
The data is about experiment to analyze the survival of lung cancer patients under the above stated two conditions. We have information about the patients, his lung cancer condition, and his survival status in days.![Chart

Description automatically generated with medium confidence]()

In this situation we want the probability of the event (death) happening to be low. The final dependent variables is a combination of Survival in days and the status (1=dead, 0=censored). Censored meaning the event didn’t happen during the period of study, the patient is alive, or it can also mean that for some reason the patient choose not to participate in the experiment. Fitting a Kaplan-Meier analysis model and with 2 groups and plotting the survival probability gives us the above plot. The survival probability is 1, 100% chance of being alive in the beginning as the days passes it drops.

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

In Treatment Group 1, where patients are only treated with the standard chemotherapy and radiation. In this case the probability that the patient will survive for 1 year (365 days) is 5.3% and for 6 months (182 Days) is 21.21%.

In Treatment Group 2, where the patients are treated with the chemotherapy and a test drug. In this case the probability that the patient will survive for 1 year (365 days) is 10.98% and for 6 months (182 Days) is 21.62%.

**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?**

The median number of days where a patient is expected to survive if they are Group 1, where patients are only treated with the standard chemotherapy and radiation is 100 days. That is the when the patient in group 1 has 50% survival chance.

The median number of days where a patient is expected to survive if they are Group 2, where the patients are treated with the chemotherapy and a test drug is 52 days. That is the when the patient in group 2 has 50% survival chance.

**Create three semi-parametric and parametric models to estimate the marginal effects of relevant predictors on survival outcomes. Interpret the coefficients of these models to explain the precise effects of age and months of diagnosis on survival probabilities of patients with standard and test treatments.**

**Models:** Cox proportional Semi parametric model, Exponential and Weibull parametric models

**lcan.cox2** <- coxph(y ~ treatment + cell.type + months.from.diagnosis + age + treatment\*age + treatment\*months.from.diagnosis + prior.chemotherapy)

**lcan.exp2** <- survreg(y ~ treatment + cell.type + months.from.diagnosis + age + treatment\*age + treatment\*months.from.diagnosis + prior.chemotherapy, dist="exponential")

**lcan.weibull2** <- survreg(y ~ treatment + cell.type + months.from.diagnosis + age + treatment\*age + treatment\*months.from.diagnosis + prior.chemotherapy, dist="weibull")

**Effect of age and months from diagnosis on survival in Treatment Group 1**

**Using Cox Semi parametric model:** People with higher age have higher risk of death with lung cancer. As the age by 10 the risk of death increases by 5%. People who are diagnosed with cancer earlier have higher risk of death. From the observation period, 10 months earlier they are diagnosed the risk of death increases by 13%.

**Using exponential parametric model:** People with higher age have higher risk of death with lung cancer. As the age by 10 the survival chance decreases by 7%. People who are diagnosed with cancer earlier have higher risk of death. From the observation period, 10 months earlier they are diagnosed the survival chance decreases by 12.5%

**Using Weibull parametric model:** People with higher age have higher risk of death with lung cancer. As the age by 10 the survival chance decreases by 7.4%. People who are diagnosed with cancer earlier have higher risk of death. From the observation period, 10 months earlier they are diagnosed the survival chance decreases by 12.3%

**Effect of age and months from diagnosis on survival in Treatment Group 2**

**Using Cox Semi parametric model:** People with higher age have higher risk of death with lung cancer. As the age by 10 the risk of death increases by 3.94%. People who are diagnosed with cancer earlier have higher risk of death. From the observation period, 10 months earlier they are diagnosed the risk of death increases by 9.3%.

**Using exponential parametric model:** People with higher age have higher risk of death with lung cancer. As the age by 10 the survival chance decreases by 5.73%. People who are diagnosed with cancer earlier have higher risk of death. From the observation period, 10 months earlier they are diagnosed the survival chance decreases by 9.92%

**Using Weibull parametric model:** People with higher age have higher risk of death with lung cancer. As the age by 10 the survival chance decreases by 5.63%. People who are diagnosed with cancer earlier have higher risk of death. From the observation period, 10 months earlier they are diagnosed the survival chance decreases by 9.85%

**Inferences:**

**Looking at the above models, we can infer than the drug has a positive effect on prolonging the survival of the patient, more than any immediate results the drug has a positive effect in the long run. From 50 to 125 days of survival the survival without the drug is more, later the survival probability is more. We can infer that the drug makes the person weak during its effect on the body, and if the individual can cope up that he can survive for a much longer time. Few other inferences are cell type 1=squamous is the least harmful type of lung cancer and cell type 3=adeno is the most harmful type of lung cancer compared to others.**