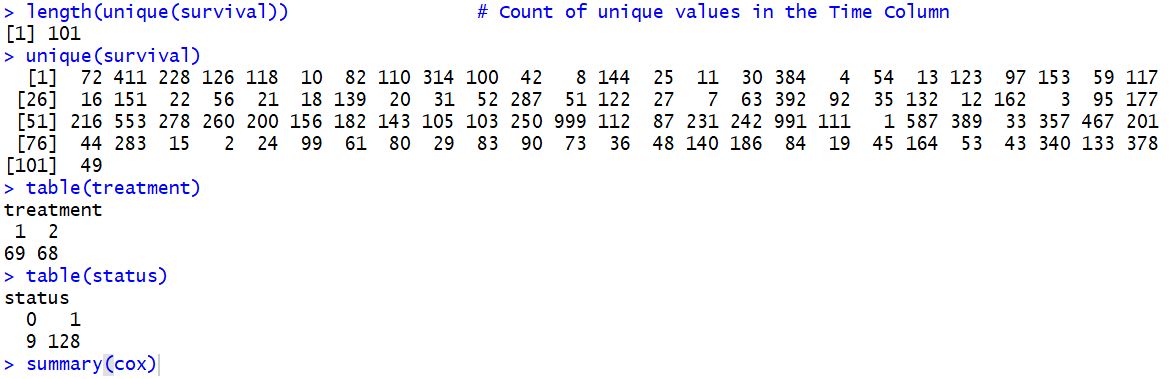
**SDM Assignment 6**

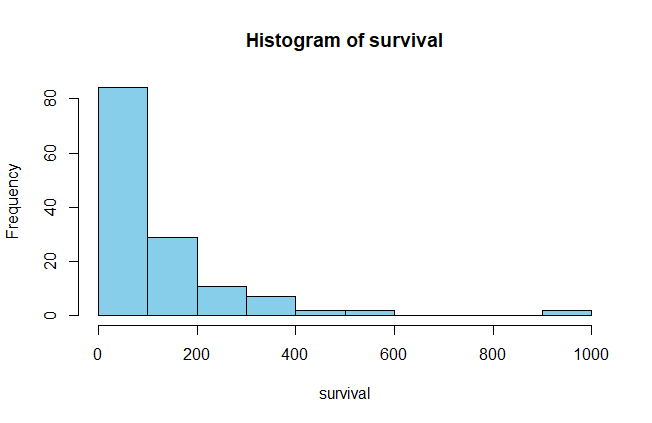
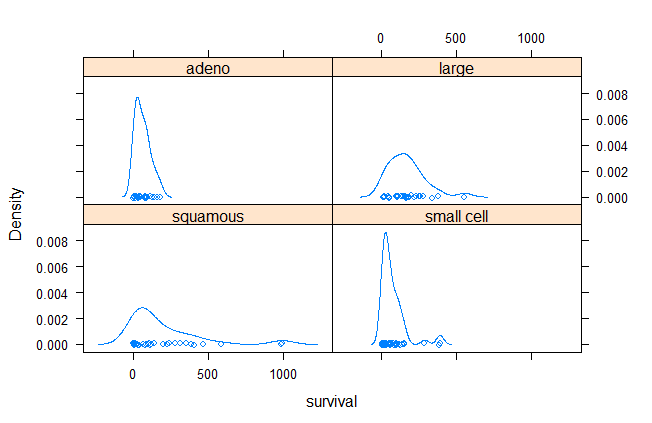
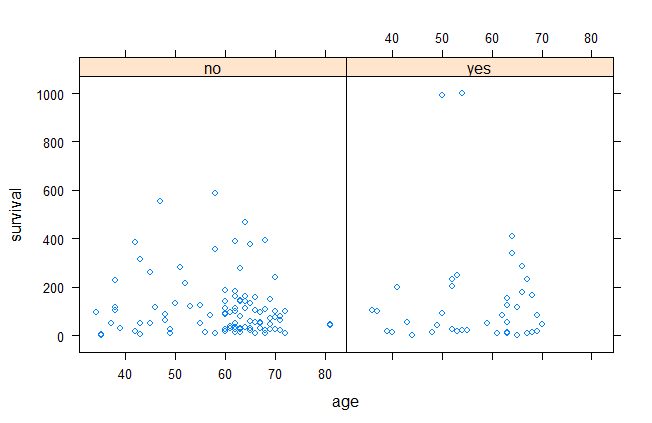
**Variables/Predictors used for the analysis**

|  |  |  |
| --- | --- | --- |
| **Predictors** | **Impact** | **Explanation/Rationale** |
| **Cell\_Type** | **+/-** | Generally Squamous cell type are tend to be found in the central part of the lungs. Whereas larger size cells appear in any part of the lungs which spreads rapidly. Hence cell type can have positive or negative effect on survival probability depending on other factors such as age, type of treatment and months after diagnosis. |
| **Treatment** | +/- | Type of treatment can have positive or negative impact on the patient’s health. Patients treated with new drug may recover from the disease early but as it is a test treatment it can also impact negatively. |
| **Karnofsky Score** | + | Karnofsky Score is the general ability of cancer patients to perform ordinary task. Maximum range of this score is 100 which depict that patient can survive longer with increase in Karnofsky Score |
| **Log(Months\_from\_diagnosis)** | **+/-** | More number of months since diagnosis may not impact patient if he has good Karnofssky score then but if the patient is at risk then his chance of surviving might decrease with increase in months\_from\_diagnosis. |
| **Age** | **-** | Generally old aged people face more risk when it comes to survival. |
| **Prior\_Chemotherapy** | **+** | A patient who have done Chemotherapy may show some quick positive impact with the treatment thus increasing the probability. |

**Exploratory Data Analysis.**

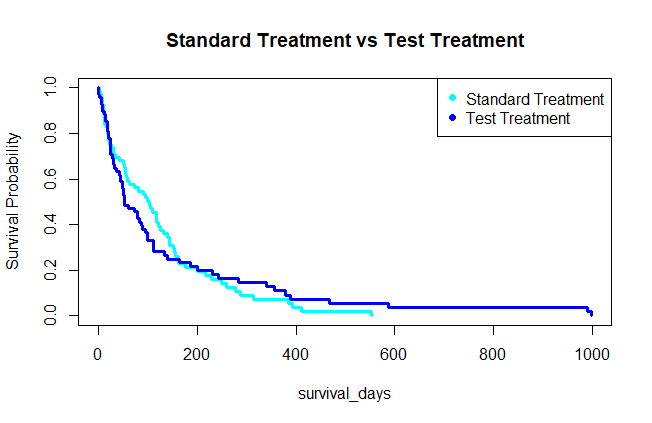
****

In the data we have **101 unique survival values** (in days). Moreover, from the data we can infer that out of **137 patients 69 got standard treatment** and **68 got test treatment** along with new drug. Out of 137 patients **128 patients** has status 1 which implies that they **are dead** whereas **9 patients are censored**.

Distribution of Survival in days Density of survival by cell type Density of Survival by age

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



From the Kaplan-Meier survival graph we can infer that patients getting standard treatment are more likely to die early than people getting test treatment.

* **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?**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **1 year or 365 days** | **6 months or 183 days** | **Interpretation** |
| **Standard Treatment** | **0.0708** | **0.2124** | For the patient undergoing standard treatment there is **7.08% chance that he will survive for 1 year** and there is **21.24% chance that the patient will survive for 6 months**. |
| **Test Treatment** | **0.1098** | **0.2329** | For the patient undergoing standard treatment there is **10.98% chance that he will survive for 1 year** and there is **23.29% chance that the patient will survive for 6 months**. |

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

|  |  |  |
| --- | --- | --- |
|  | **Mean Number of Days** | **Interpretation** |
| **Standard Treatment** | **100 days** | For the patient undergoing standard treatment mean number of days he can be expected to survive is **100 days**. |
| **Test Treatment** | **52 days** | For the patient undergoing test treatment mean number of days he can be expected to survive is **52 days**. |

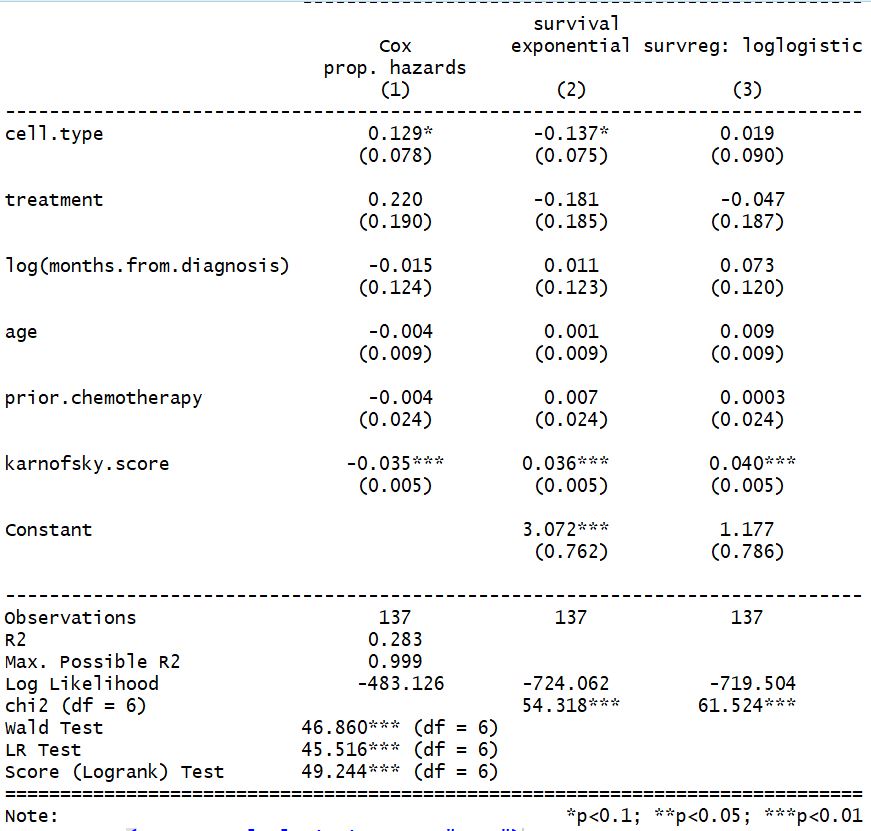
2. **Create three semi-parametric and parametric models to estimate the marginal effects of relevant predictors on survival outcomes. Interpret the coefficients of these models to explain the precise effects of age and months of diagnosis on survival probabilities of patients with standard and test treatments.**

In order to estimate survival I ran 3 models, 1 semi parametric and 2 parametric models for exponential and loglogistic distributions. I have interpreted the cox proportional model in order to present the insights.

cox <- coxph(Surv(survival, status) ~ cell.type + treatment + log(months.from.diagnosis) + age + prior.chemotherapy + karnofsky.score, method="breslow")

exp <- survreg(Surv(survival, status) ~ cell.type + treatment + log(months.from.diagnosis) + age + prior.chemotherapy + karnofsky.score, dist="exponential")

loglogistic <- survreg(Surv(survival, status) ~ cell.type + treatment + log(months.from.diagnosis) + age + prior.chemotherapy + karnofsky.score, dist="loglogistic")



**Interpretations**

1. With change in the **cell type 1** the **probability of patient dying increases by 13.76%**. For **cell type 2** the **probability of patient dying increases by 29.42%.** For **cell type 3** the **probability of patient dying increases by 47.23%**. For **cell type 4** the **probability of patient dying increases by 67.5%.**
2. For **standard treatment** the **probability of patient dying increases by 24.58%** and for **test treatment** the **probability of dying increases by 55.2%.**
3. With **1 month increase in log of months** from diagnosis the **probability of patient dying decreases by 1.47%.**
4. With **1 year increase in the patient’s a**ge the **chance of patient dying goes down by 0.38%.**
5. If **patient has done chemotherapy prior** then his **chance of dying goes down by 0.4%.**
6. **I unit increase in Karnofsky score** **drops the chance of patient dying by 3.5%.**