**Dominic Boccaleri**

**Final Project**

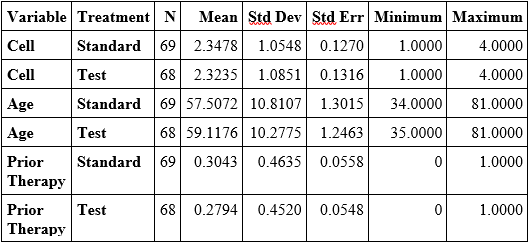
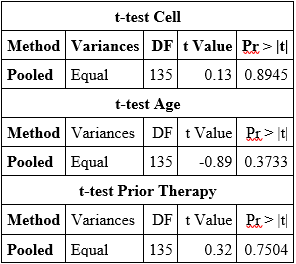
**Background:**

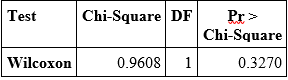
Lung cancer is a malignant lung tumor that is characterized as uncontrolled cell growth in the tissues of the lung. In general, there are two types of treatment, surgery followed by chemotherapy and strictly chemotherapy, where the tumor is inoperable. In a study of 322 elderly patients, older than 70, two methods of chemotherapy were given. Patients who received a more “platinum-based” treatment were more likely to have a better median survival time and a 1-year survival with a p < 0.001 (Yau). Improvements of chemotherapy can increase patient satisfaction and increase system performance. This is especially important for the incurable lung cancer patients (von Plessen). “Small-cell **lung cancer** (SCLC) is thought to be sensitive to **chemotherapy**; therefore, second-line **chemotherapy** is recommended. Although platinum rechallenge is performed in the second-line **chemotherapy** for sensitive-relapsed SCLC, it remains unclear whether such a strategy is effective” (Shiozawa). This study is important because lung cancer generally has a poor prognosis and remains relatively lethal compared to prior survival (Gebbia). Inoperable lung cancer’s only treatments are different types of chemotherapy and finding one that is most effective gives patients the best chance of survival. This study follows VA patients with inoperable lung cancer so finding ways to treat this disease is imperative for their survival in the future. With the data, we want to find if one of the treatment groups has a better survival time than then other. Also, we want to see if the age of the patient increases the risk of dying due to having inoperable lung cancer.

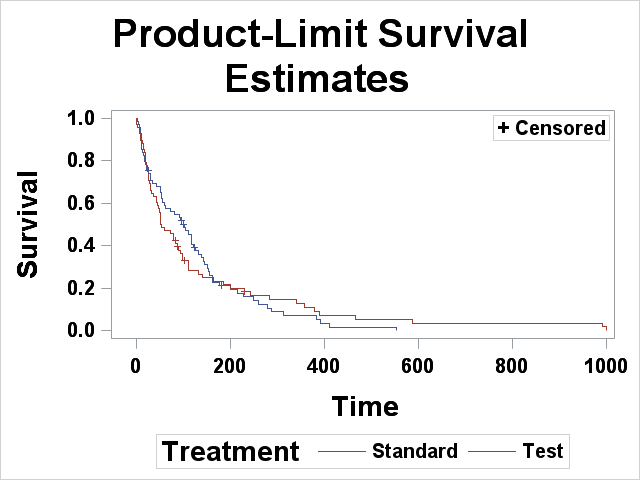
**Description of the Statistical Methods:**

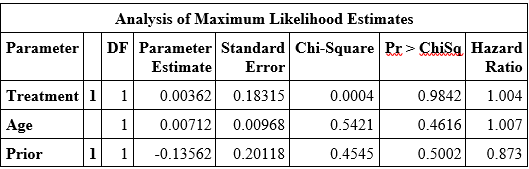
When analyzing survival times of cancer patients, we will use a non-parametric Wilcoxon chi-square test to compare the treatment groups. We split the sample into two groups, those who were administered the standard chemotherapy treatment and test/experimental chemotherapy treatment. We look at when the person experienced the event or was censored. The survival time is given in days between 1 and 999 days. The null hypothesis is given as the survival time of the standard treatment group is the same as the survival time for the test chemotherapy group. In the comparing of survival times, we ran the Tarone-Ware: Wilcoxon test on the survival times. This is because it should have the best p-value due to the fact that the survival curves cross based on the data. In order to perform the analysis we need assume that the variable we split the data on is categorical and that there are at least five successes and five failures for the test.

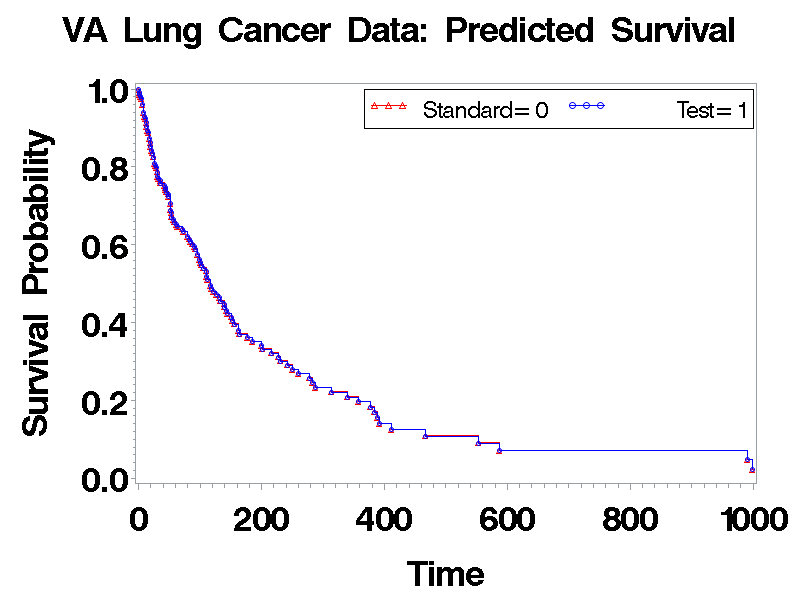
In the data we test to see if age has a significant effect on the distribution of survival times of VA lung cancer patients. Age ranges from 34 to 81 years old. The null hypothesis states that age has no significant effect on the survival times between the standard chemotherapy group and the test group. The alternate hypothesis states that there is a significant difference in the survival times between the standard and test treatment groups as a result of the age of the patient. We will use the Cox Proportional Hazard model to analyze the data. The proportional hazards model assumes a continuous hazard over the entire time and that there are no ties between data points. If there are ties, you need to use a modification to adjust the likelihood for ties between survival time data. The necessary assumptions are met and any ties will be broken by the model.

**Statistical Results:**

As seen above, the variables are broken down by treatment. The mean and standard deviations are relatively similar. Each of the variances among the three variables are equal so the Pooled method was used. The Cell variable is a p-value of 0.8945, the Age variable has a p-value of 0.3733, and the Prior Therapy variable has a p-value of 0.7504 in the 2 sample t-test. This shows that with the data of each variable, split by treatment type, we do not have significant evidence to show that the treatment groups are different among Cell Type, Age, or Prior Therapy. Variables like “how long someone has had cancer before treatment” are lurking variables which may affect the study.

 Based on the non-parametric approach, the Wilcoxon method produced a p-value of 0.3270. From this, we fail to reject the null hypothesis. In conclusion, we do not have significant evidence to prove that the survival times of VA lung cancer patient is different depending on the type of treatment they received.

 From the semi-parametric method, we use the Cox Proportional Hazard (CPH) model to calculate the ratio of the risks and if a lung cancer patient’s age has a significant effect on the survival times between the two treatment groups. After calculating the CPH, we see that Age has a p-value of 0.4616. This means that we should fail to reject the null hypothesis and we cannot say with evidence that age has significant effect on the survival time between the standard treatment group and the test group. If using the model, we would say that parameter estimate of age is positive, so as age increases, so does the risk of dying from lung cancer as a VA lung cancer patient. Also, the coefficient of age is 0.00712 so for each additional year in the age of a person, the risk increases by times relatively to the baseline risk. Treatment type has a positive coefficient the so those using the test treatment actually experience an increase risk over the standard group. The hazard ratio is 1.004 which means that the risk is increased by 1.004 times the risk of the standard treatment. Prior Therapy has a negative coefficient so the patients who have had prior therapy for inoperable lung cancer experience a decreased risk of experiencing the death event. Prior Therapy has a 0.873 hazard ratio meaning that those who have had prior therapy have a reduction of 12.7% in risk over those who have not had any prior therapy. Shown below is the baseline risk associated with each treatment.



The results of this study cannot be generalized without the worry of the effect of extrapolation. This is due to there being high p-values associated with data, making the results “not statistically significant” and therefore should not be extended to any other populations.

Lurking variables could throw off the results of the study. One potential lurking variable is “time from diagnosis” which could tell how much cancer there is. Another lurking variable could be “stage of the cancer”. Cancer is classified as stage 1-4 where the higher the number, the further the cancer has grown and spread. 1 being cancer is found in a lung but not spread and 4 being the cancer has spread to both lungs and the surrounding areas or distant organs. If the cancer has spread, it may be more difficult to treat since it is no longer in a specific area. Also, the size of the tumor may be a lurking variable since a smaller tumor may be able to be controlled more easily with chemotherapy.

**Report Summary:**

In the analysis, we wanted to see if one of the treatment groups had a better survival time distribution than the other and if the age of a person had an effect on the survival of a lung cancer patient. The data used came from VA (veteran) patients with inoperable lung cancer. Variables used were survival time in days, status which was censored or not, treatment which was standard or the experimental test treatment, age of the patient in years, the cell type which was what type of cancer the patient had, and whether or not the patient had had prior therapy before the treatment. Two analyses were done, non-parametric Tarone-Ware: Wilcoxon test and semi-parametric Cox Proportional Hazard model. The Wilcoxon test compared the survival times of the two treatment groups to decide if patients using one of the treatments had greater probability of experiencing death at a later time. The Cox Proportional Hazard model allowed for the comparison of the two groups to see if age had a significant effect on the distribution of survival times. Each of the necessary assumptions were met. For the non-parametric model, we had at least 5 successes and failures and that the data was split on a categorical variable, treatment type in this case. In the case of the semi-parametric model, we could not have any ties in the data, or we had to use a method of breaking these ties, and continuous hazard across the entire distributions. The p-values of the 2 sample t-test were all greater than 0.05 so there were no difference in the means of the data divided by the treatment group. From the Wilcoxon test, we saw a p-value of 0.3270 which allowed us to fail to reject the null hypothesis and say that between the two treatment groups, we did not have significant evidence to conclude that the survival times were different. In the Cox Proportional Hazard model, the p-value for age was 0.4616 so we failed to reject the null hypothesis and said that there wasn’t enough evidence to conclude that age significantly affected the distribution of survival times between the two treatment groups. The coefficient of age and treatment variables are positive so as age increases, so does the risk associated with it, and if a patient is administered the test treatment, the risk is also increased. The coefficient of prior therapy is negative so if the patient has had prior therapy, the risk is decreased. The hazard ratio of treatment type and age are 1.004 and 1.007, respectively, so if the person gets the test treatment, their risk is increased by 1.004 times that of a patient who takes the standard treatment. Also, for each additional year in a patient’s age, the risk associated increases by 1.007 times relative to the baseline risk. The hazard ratio of prior therapy is 0.873 so patients who have had prior therapy saw a 12.3% reduction in risk compared to those who have not had any prior therapy. A follow up study should follow patients based on the size of the cancer and the stage of the cancer to better account for patients who may be harder to treat due to the growth of the cancer.

**Bibliography:**

Gebbia, Vittorio, et al. “Evolution of Therapy Decision-Making Process for Advanced Non-Small Cell Lung Cancer.” *Oncology*, vol. 77, Dec. 2009, pp. 97–102. *EBSCOhost*, doi:10.1159/000258501.

Shiozawa, Toshihiro, et al. “Rechallenge with First-Line Platinum Chemotherapy for Sensitive-Relapsed Small-Cell Lung Cancer.” *Case Reports in Oncology*, vol. 11, no. 3, Sept. 2018, pp. 622–632. *EBSCOhost*, doi:10.1159/000492780.

von Plessen, Christian. “Improving Chemotherapy for Patients with Advanced Non-Small Cell Lung Cancer.” *Clinical Respiratory Journal*, vol. 5, no. 1, Jan. 2011, pp. 60–61. *EBSCOhost*, doi:10.1111/j.1752-699X.2010.00199.x.

Yau, T., et al. “Time and Chemotherapy Treatment Trends in the Treatment of Elderly Patients (Age 70 Years) with Small Cell Lung Cancer.” *British Journal of Cancer*, vol. 94, no. 1, Jan. 2006, pp. 18–21. *EBSCOhost*, doi:10.1038/sj.bjc.6602888.