Survival Analysis of Lung Cancer Patients

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

**a. 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?**

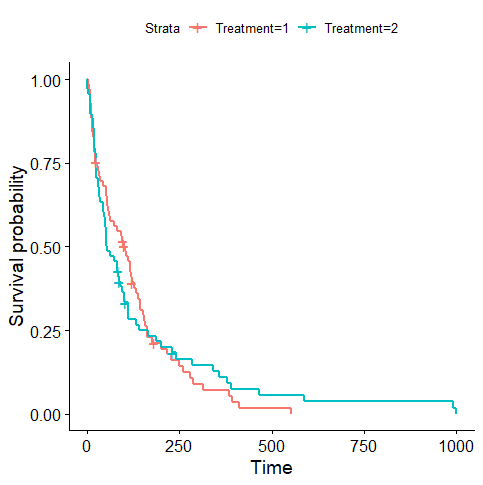
|  |  |  |
| --- | --- | --- |
|  | Treatment 1 | Treatment 2 |
| Survival Probability (365 days) | ~0.06 | ~0.10 |
| Survival Probability (183 days) | ~0.20 | ~0.21 |

#1 year (365 days)- the survival probability of patient with standard treatment is approximately ~0.06 whereas the survival probability of patients with Test drug is 0.10 which shows positive effectiveness of Drug under test after a year (over a long period).

# 6 months (183 days), the survival probability of patient with standard treatment is approximately ~0.20 and the survival probability of patients with Test drug is 0.21 approximately equal, which shows Drug under test has no significant effects on patients in 6 months (short term).

**#b. 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?**

#At Survival probability 0.50, i.e., patient under standard treatment are expected to survive for 100 days, and patients under test treatment are expected to be about 52 days. Test treatment is ineffective.



**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.**

**#Semi Parametric Model--- Cox proportional hazard model - coefficients and hazard rates**

**Interpretation-**

-Patients with **test treatment** will more likely **die soon and** increases hazard rate by 33%

-Patients with **Cell Type "3"/ adeno** are more likely to **die soon**, with an increase in hazard rate by 228% followed by **cell Type 2/** **small cell, Cell Type 4/** **large**, with an increase in hazard rate by 135% and 49% respectively when compared to **cell type "1"/** **squamous**. Adeno is deadly type of Cell.

-Patients with more **months of Diagnosis** has approximately **no effect** on death.

-Increase in **Age** has approximately **no effect** on death or (little evidence to tell elder people die soon)

-Patients with increasing **karnofsky\_score** is likely to **live more** and with hazard rate decrease by 4%. Healthy people live longer with cancer.

-Patients with **Prior Chemotherapy** will die soon and increases hazard rate by 7%. Doctors/researchers need to check clinically with more experiment’s effects of Chemotherapy on cancer.

**# Exponential, Weibull, and log-logistic parametric model coefficients**

**Interpretation**

From the above Summary, parametric models also suggest that

* **Treatment** 2(test) decreases time to death by a factor of 0.22(Exponential model), 0.23(Weibull model) 0.08 (Loglogistic model) i.e., **test treatment is not preferred type of treatment.**
* **Age** increases time to death by a factor of 0.006(Exponential model), 0.006(Weibull model) 0.009 (Loglogistic model)., i.e. very negligible which suggests that **age has no effect** on time to death and treatment type.
* **Karnofsky score**- with increase of this score by 1 point the time to death increase by a factor of 0.03(Exponential model), 0.03(Weibull model) 0.03 (Loglogistic model). Which suggest **healthy people will survive** for more days.
* **Months from Diagnosis-** increases time to death by a factor of 0.0003(Exponential model), 0.0005(Weibull model) 0.002 (Loglogistic model)., i.e. very negligible which suggests that **months from diagnosis has no effect** on time to death and treatment type.
* **Prior Chemotherapy-** If a patient is treated early by chemotherapy, time to death is decreased by a factor of 0.05, (Exponential model),0.04 (Weibull model) 0.1 (Loglogistic model).
* **Cell type - P**atients with Cell type(3) decrease time to death by a factor of 1.11, (Exponential model),1.13 (Weibull model) 0.74 (Loglogistic model) followed by cell type(2) 0.82(Exponential model),0.82(Weibull model),0.7(Loglogistic model) cell type 4 0.38(Exponential model),0.40(Weibull model),(-0.017(Loglogistic model)~negligible) compared to cell type 1.

**Cell type 3 is more deadly than type 2 followed type 4 and type1.**

**R code and Summaries**

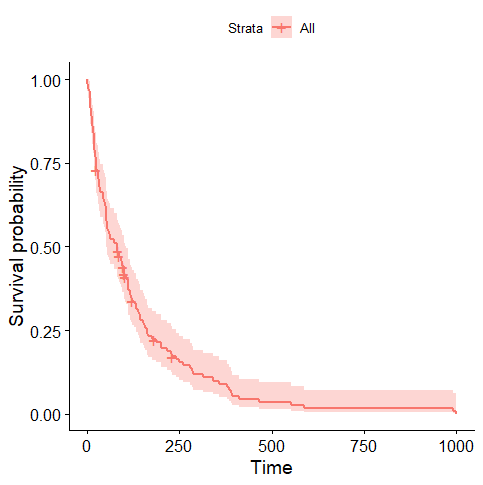
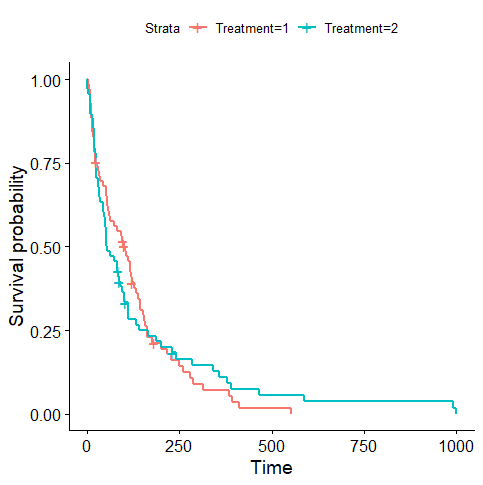
# Kaplan-Meier non-parametric analysis

# Group data based on Time and estimate KM survival function based on Event

km1 <- survfit(Surv(Survival\_in\_days, Status) ~ 1)

summary(km1)

ggsurvplot(km1, data = d, pval = TRUE)

#From the Plot we can see Survival probability decreases as the days passes for a Cancer Patient

# Kaplan-Meier non-parametric analysis by group

km2 <- survfit(Surv(Survival\_in\_days, Status) ~ Treatment)

summary(km2)

ggsurvplot(km2, data = d, pval = TRUE)

#From the above graph we can see that with test treatment the survival probability is more in a long run, which means the test is giving positive results on lung cancer patients only after a certain period of time.

# Cox proportional hazard model - coefficients and hazard rates

cox <- coxph(Surv(Survival\_in\_days, Status) ~ as.factor(Treatment) + Age + as.factor(Cell\_type) + Karnofsky\_score + Months\_from\_Diagnosis + as.factor(Prior\_chemotherapy), method="breslow")

summary(cox)

|  |
| --- |
| summary(cox)  coef exp(coef) se(coef) z Pr(>|z|)  as.factor(Treatment)2 0.289936 1.336342 0.207210 1.399 0.16174  Age -0.008549 0.991487 0.009304 -0.919 0.35816  as.factor(Cell\_type)2 0.856487 2.354873 0.275190 3.112 0.00186  as.factor(Cell\_type)3 1.188299 3.281496 0.300763 3.951 7.78e-05  as.factor(Cell\_type)4 0.399628 1.491270 0.282663 1.414 0.15742  Karnofsky\_score -0.032622 0.967905 0.005505 -5.926 3.11e-09  Months\_from\_Diagnosis -0.000092 0.999908 0.009125 -0.010 0.99196  as.factor(Prior\_chemotherapy)10 0.072327 1.075006 0.232133 0.312 0.75536    as.factor(Treatment)2  Age  as.factor(Cell\_type)2 \*\*  as.factor(Cell\_type)3 \*\*\*  as.factor(Cell\_type)4  Karnofsky\_score \*\*\*  Months\_from\_Diagnosis  as.factor(Prior\_chemotherapy)10  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  exp(coef) exp(-coef) lower .95 upper .95  as.factor(Treatment)2 1.3363 0.7483 0.8903 2.0058  Age 0.9915 1.0086 0.9736 1.0097  as.factor(Cell\_type)2 2.3549 0.4247 1.3732 4.0384  as.factor(Cell\_type)3 3.2815 0.3047 1.8200 5.9167  as.factor(Cell\_type)4 1.4913 0.6706 0.8569 2.5951  Karnofsky\_score 0.9679 1.0332 0.9575 0.9784  Months\_from\_Diagnosis 0.9999 1.0001 0.9822 1.0180  as.factor(Prior\_chemotherapy)10 1.0750 0.9302 0.6821 1.6943  Concordance= 0.736 (se = 0.021 )  Likelihood ratio test= 61.41 on 8 df, p=2e-10  Wald test = 61.65 on 8 df, p=2e-10  Score (logrank) test = 65.92 on 8 df, p=3e-11 |
|  |
| |  | | --- | |  | |

# Exponential, Weibull, and log-logistic parametric model coefficients

exp <- survreg(Surv(Survival\_in\_days, Status) ~ as.factor(Treatment) + Age + as.factor(Cell\_type) + Karnofsky\_score + Months\_from\_Diagnosis + as.factor(Prior\_chemotherapy), dist="exponential")

summary(exp)

weibull <- survreg(Surv(Survival\_in\_days, Status) ~ as.factor(Treatment) + Age + as.factor(Cell\_type) + Karnofsky\_score + Months\_from\_Diagnosis + as.factor(Prior\_chemotherapy), dist="weibull")

summary(weibull)

loglogistic <- survreg(Surv(Survival\_in\_days, Status) ~ as.factor(Treatment) + Age + as.factor(Cell\_type) + Karnofsky\_score + Months\_from\_Diagnosis + as.factor(Prior\_chemotherapy), dist="loglogistic")

summary(loglogistic)

library(stargazer)

stargazer(cox, exp, weibull, loglogistic, type="text")

|  |
| --- |
| stargazer(cox, exp, weibull, loglogistic, type="text")  =============================================================================================  Dependent variable:  -------------------------------------------------------------  Survival\_in\_days  Cox exponential Weibull survreg: loglogistic  prop. hazards  (1) (2) (3) (4)  ---------------------------------------------------------------------------------------------  as.factor(Treatment)2 0.290 -0.220 -0.229 -0.088  (0.207) (0.199) (0.187) (0.179)    Age -0.009 0.006 0.006 0.009  (0.009) (0.009) (0.009) (0.009)    as.factor(Cell\_type)2 0.856\*\*\* -0.820\*\*\* -0.826\*\*\* -0.708\*\*\*  (0.275) (0.262) (0.246) (0.249)    as.factor(Cell\_type)3 1.188\*\*\* -1.113\*\*\* -1.133\*\*\* -0.743\*\*\*  (0.301) (0.276) (0.258) (0.272)    as.factor(Cell\_type)4 0.400 -0.377 -0.398 0.017  (0.283) (0.273) (0.255) (0.270)    Karnofsky\_score -0.033\*\*\* 0.031\*\*\* 0.030\*\*\* 0.036\*\*\*  (0.006) (0.005) (0.005) (0.004)    Months\_from\_Diagnosis -0.0001 -0.0003 -0.0005 0.002  (0.009) (0.009) (0.008) (0.010)    as.factor(Prior\_chemotherapy)10 0.072 -0.049 -0.044 -0.102  (0.232) (0.227) (0.212) (0.211)    Constant 3.189\*\*\* 3.262\*\*\* 2.024\*\*\*  (0.704) (0.663) (0.678)    ---------------------------------------------------------------------------------------------  Observations 137 137 137 137  R2 0.361  Max. Possible R2 0.999  Log Likelihood -475.179 -716.159 -715.551 -711.943  chi2 (df = 8) 70.124\*\*\* 65.080\*\*\* 76.646\*\*\*  Wald Test 61.650\*\*\* (df = 8)  LR Test 61.409\*\*\* (df = 8)  Score (Logrank) Test 65.917\*\*\* (df = 8)  =============================================================================================  Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01 |
|  |
| |  | | --- | | > | |