Assignment -3

Statistical Modeling (Survival Analysis)

13/05/2024

Anik Saha

Introduction:

Survival analysis is a statistical method used to analyze the time until an event of interest occurs. In this report, we analyze survival data to understand the factors influencing survival rates and to compare survival between different groups.

About my data:

Cervical cancer data

Sample of 30 patients from a randomized study of radiotherapy with and

without a new radiosensitiser. Based on data from MRC Working Party on

Advanced Carcinoma of the Cervix, Radiotherapy Oncology 26:93-103, 1993.

Analyzed in and obtained from MKB Parmar, D Machin, Survival Analysis: A

Practical Approach, Wiley, 1995. Group = treatment (2 = radiosensitiser),

age = age in years at diagnosis, status: (0 = censored)

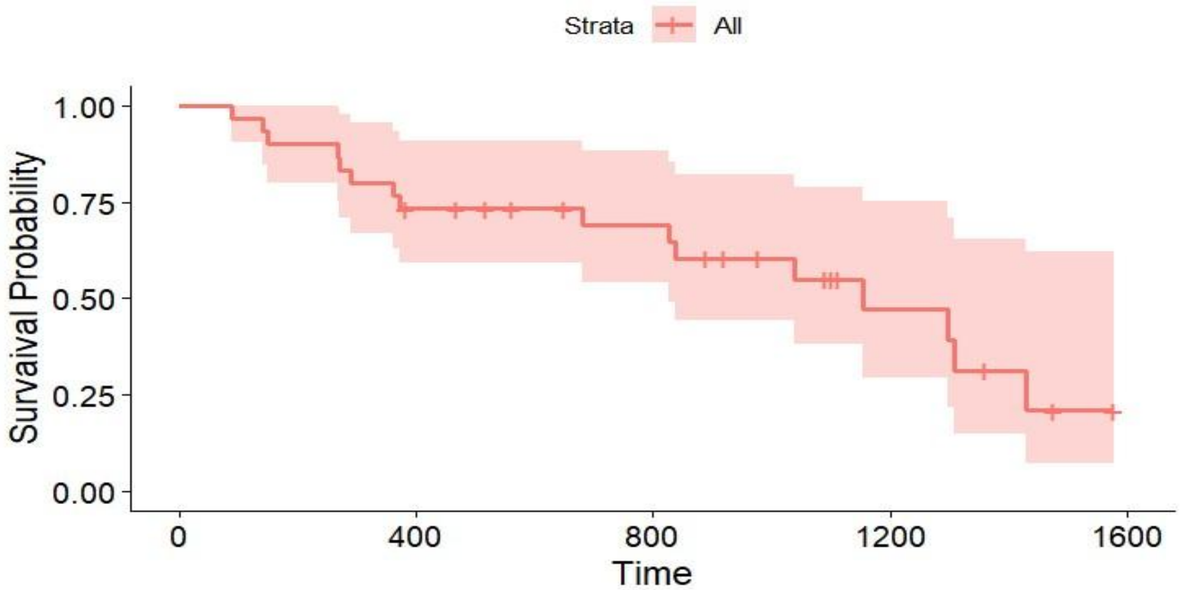
Survival time is in days (from randomization).

Summary of my data:

```
    Group            Age           Survival          Status      
 Min.   :1.000   Min.   :27.00   Min.   :  90.0   Min.   :0.0000  
 1st Qu.:1.000   1st Qu.:45.75   1st Qu.: 375.5   1st Qu.:0.0000  
 Median :1.000   Median :57.00   Median : 832.0   Median :1.0000  
 Mean   :1.467   Mean   :55.43   Mean   : 786.7   Mean   :0.5333  
 3rd Qu.:2.000   3rd Qu.:63.75   3rd Qu.:1109.8   3rd Qu.:1.0000  
 Max.   :2.000   Max.   :74.00   Max.   :1577.0   Max.   :1.0000  
```

Methods:

We first performed a Kaplan-Meier estimation to visualize survival probabilities over time and compare survival between groups. Subsequently, we conducted a log-rank test to formally assess the difference in survival between groups. Finally, we fitted a Cox proportional hazards model to identify factors influencing survival rates.



z

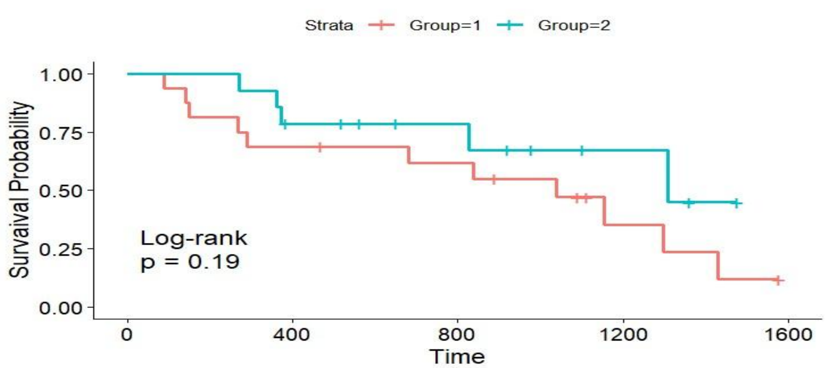
(Kaplan – Meier Curve)

Results:

Kaplan-Meier Estimate:

The Kaplan-Meier estimate revealed similar survival probabilities.

Log-Rank Test:



The log-rank test did not show a significant difference in survival between groups, suggesting that the null hypothesis of equal survival rates for both groups cannot be rejected.

Cox Proportional Hazards Model:

```
fit=coxph(formula = Surv(Survival,Status)~Age+Group,data = a)
> summary(fit)
Call:
coxph(formula = Surv(Survival, Status) ~ Age + Group, data = a)
```

```
  n= 30, number of events= 16

          coef exp(coef) se(coef)      z Pr(>|z|)
Age   -0.02162   0.97861  0.02343 -0.923    0.356
Group -0.68344   0.50488  0.54598 -1.252    0.211

      exp(coef) exp(-coef) lower .95 upper .95
Age      0.9786      1.022    0.9347     1.025
Group    0.5049      1.981    0.1732     1.472

Concordance= 0.612  (se = 0.068 )
Likelihood ratio test= 2.57  on 2 df,   p=0.3
Wald test            = 2.33  on 2 df,   p=0.3
Score (logrank) test = 2.43  on 2 df,   p=0.3
```

The Cox proportional hazards model indicated that neither Age nor Group had a significant impact on survival rates, as evidenced by the non-significant p-values.

Conclusion:

Based on my analysis, I conclude that there is no evidence to suggest a difference in survival between the groups studied. Age and Group do not appear to be significant factors affecting survival rates in this dataset.

R-Code:

```
a=read.csv("survival data.csv")

summary(a)

# Kaplan-Meier estimate

library(survival)

km=survfit(formula = Surv(Survival,Status)~1,data = a)

summary(km)

library(ggplot2)

library(survminer)

ggsurvplot(km,xlab = "Time",ylab = "Survaival Probability",main="Kaplan-Meier estimate")

#log rank test

log_rank=survfit(Surv(Survival,Status)~Group,data = a)

summary(log_rank)

#plots

ggsurvplot(log_rank,pval=T,pval.method = T,xlab="Time",ylab="Survaival Probability")

# cox proportional hazard model

fit=coxph(formula = Surv(Survival,Status)~Age+Group,data = a)

summary(fit)
```