Leukemia patients - Survival Analysis

# Description:

This report analyzes the effect of a treatment on 42 children with acute leukemia . Patients were observed until they relapsed ,dropped out ,or the study ended.

# Methodology:

This report uses non-parametric techniques to observe difference between the placebo and the treatment group. A Kaplan-Meier Estimates is computed for each group in order to observe the survival function of each independently . A log rank test is then performed to test whether both groups are significantly different or not.

## Variable Descriptions

* Time : *Represents the time to relapse/censor for each patient*
* Treatment : *Represents whether subject received a treatment or not (treated=1, placebo=0 )*
* Failure : *Represents whether subject relapsed or censored (relapsed =1 , censored = 0 )*

## Imports

library(readxl)  
library(ggplot2)  
library(dplyr)  
library(epiDisplay)  
library(epicalc)  
library(survival)  
library(survminer)

df <- read\_excel("leukemia.xlsx") # total ob.servations 42

## New names:  
## • `` -> `...4`  
## • `` -> `...5`  
## • `` -> `...6`  
## • `` -> `...7`

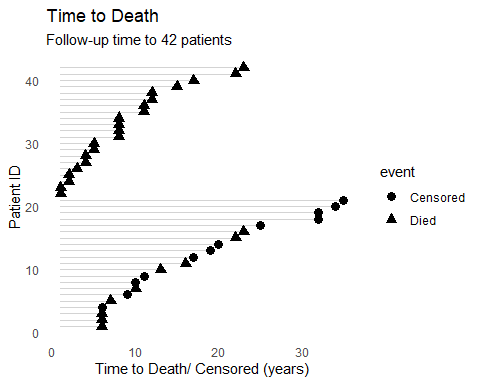
df <- df[-c(4:7)] # dropping description columns

### Wrangling

groupss <- factor(df$Treatment,labels=c("Placebo","Treatment"))  
event <- factor(df$Failure,labels=c("Censored","Died"))  
df1<- cbind(df,event)  
ID<- c(1:42)  
df1<- cbind(df,ID)  
df1$Treatment <- groupss

#### Follow-Up graph

ggplot(df1,   
 aes(x=df1$Time,   
 y=df1$ID)) +  
 geom\_segment(aes(x = 1,   
 xend = df1$Time,   
 y = df1$ID,   
 yend = df1$ID),  
 color = "lightgrey") +  
 labs (x = "Time to Death/ Censored (years)",  
 y = "Patient ID",  
 title = "Time to Death",  
 subtitle = "Follow-up time to 42 patients") +  
 theme\_minimal()+  
 theme(panel.grid.major = element\_blank(),  
 panel.grid.minor = element\_blank()) +  
 geom\_point(aes(shape=event), size = 3)



Patients 22 and above represents the placebo group

## Kaplan-Meier Estimates / Treatment group

kaplan Meir refers to…………….

df.treat <- df1[ which(df1$Treatment=='Treatment'),] # dividing into 2 groups   
  
  
KMTREAT <- survfit(Surv(df.treat$Time, df.treat$Failure) ~ 1,conf.type="log-log", data = df.treat)  
summary(KMTREAT)

## Call: survfit(formula = Surv(df.treat$Time, df.treat$Failure) ~ 1,   
## data = df.treat, conf.type = "log-log")  
##   
## time n.risk n.event survival std.err lower 95% CI upper 95% CI  
## 6 21 3 0.857 0.0764 0.620 0.952  
## 7 17 1 0.807 0.0869 0.563 0.923  
## 10 15 1 0.753 0.0963 0.503 0.889  
## 13 12 1 0.690 0.1068 0.432 0.849  
## 16 11 1 0.627 0.1141 0.368 0.805  
## 22 7 1 0.538 0.1282 0.268 0.747  
## 23 6 1 0.448 0.1346 0.188 0.680

Comment

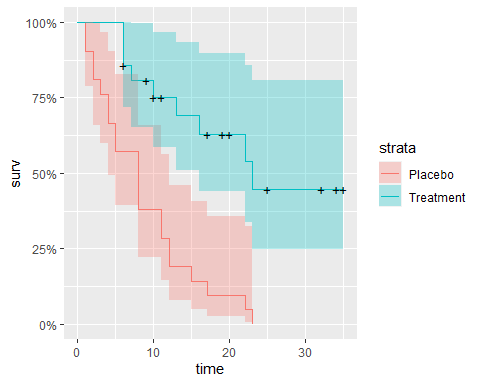
## Kaplan-Meier Estimates /placebo group

df.placebo <- df1[ which(df1$Treatment=='Placebo'),]  
  
KMPLACE <- survfit(Surv(df.placebo$Time, df.placebo$Failure) ~ 1,conf.type="log-log", data = df.placebo)  
summary(KMPLACE)

## Call: survfit(formula = Surv(df.placebo$Time, df.placebo$Failure) ~   
## 1, data = df.placebo, conf.type = "log-log")  
##   
## time n.risk n.event survival std.err lower 95% CI upper 95% CI  
## 1 21 2 0.9048 0.0641 0.67005 0.975  
## 2 19 2 0.8095 0.0857 0.56891 0.924  
## 3 17 1 0.7619 0.0929 0.51939 0.893  
## 4 16 2 0.6667 0.1029 0.42535 0.825  
## 5 14 2 0.5714 0.1080 0.33798 0.749  
## 8 12 4 0.3810 0.1060 0.18307 0.578  
## 11 8 2 0.2857 0.0986 0.11656 0.482  
## 12 6 2 0.1905 0.0857 0.05948 0.377  
## 15 4 1 0.1429 0.0764 0.03566 0.321  
## 17 3 1 0.0952 0.0641 0.01626 0.261  
## 22 2 1 0.0476 0.0465 0.00332 0.197  
## 23 1 1 0.0000 NaN NA NA

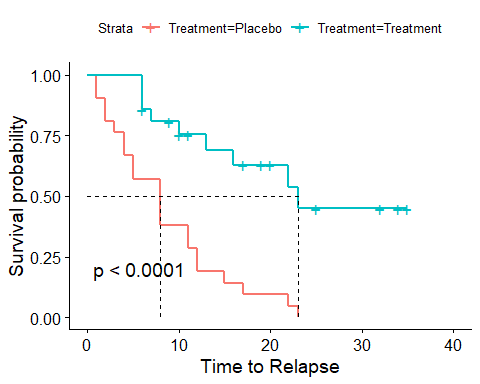
## Survival-curves

library(ggfortify)  
  
  
combinedKM <- survfit(Surv(df1$Time, df1$Failure) ~ df1$Treatment, data = df1)  
  
autoplot(combinedKM)



#### Comparing survival functions /Log Rank test

ggsurvplot(combinedKM, data = df1, pval = TRUE,   
 xlab="Time to Relapse", surv.median.line = c("hv"))



survdiff(Surv(df1$Time, df1$Failure) ~ df1$Treatment, data = df1)

## Call:  
## survdiff(formula = Surv(df1$Time, df1$Failure) ~ df1$Treatment,   
## data = df1)  
##   
## N Observed Expected (O-E)^2/E (O-E)^2/V  
## df1$Treatment=Placebo 21 21 10.7 9.77 16.8  
## df1$Treatment=Treatment 21 9 19.3 5.46 16.8  
##   
## Chisq= 16.8 on 1 degrees of freedom, p= 4e-05