

HW2 Exercise 1

Wen Fan(wf2255), Banruo Xie(bx2168), Hanjun Li(hl3339)

3/3/2020

```
library(readr)
library(ggfortify)
```

```
## Loading required package: ggplot2
```

```
library(survival)
library(ggplot2)
```

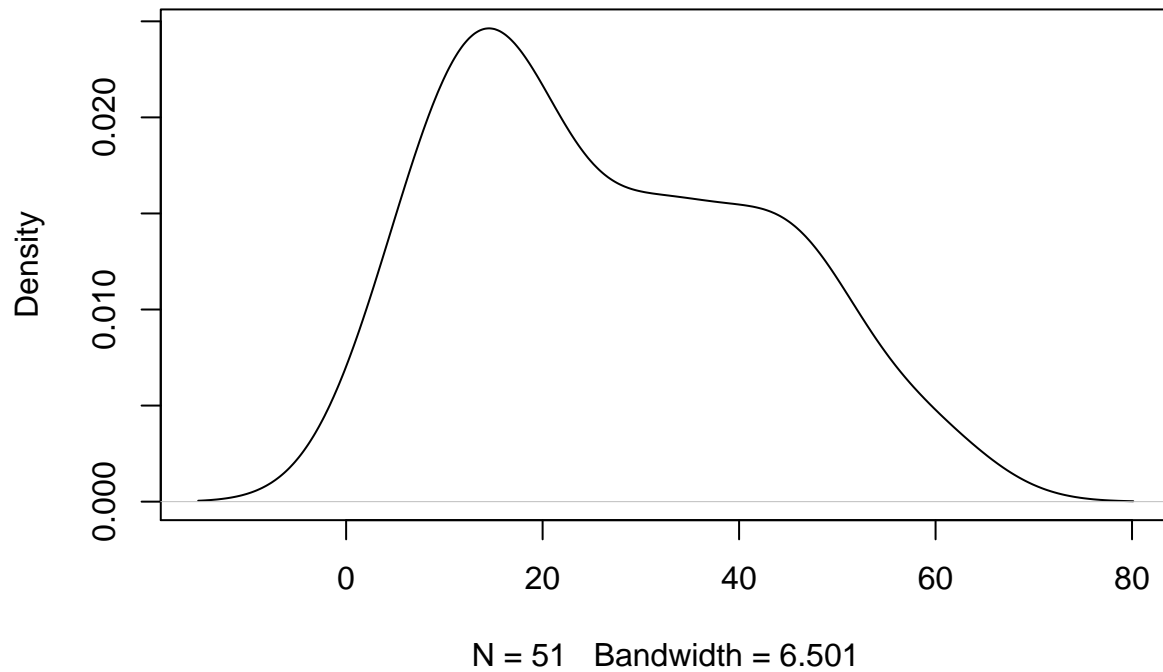
```
transplant <- read_table2("~/Documents/Columbia/STAT W5703/HW/HW2/transplant.txt",
  col_names = c("time", "type", "Indicator"), skip = 8)
```

```
## Parsed with column specification:
## cols(
##   time = col_double(),
##   type = col_double(),
##   Indicator = col_double()
## )
```

Problem 1

```
# make the variable "type" a factor
transplant$type <- factor(transplant$type)
censored_index <- which(transplant$Indicator==0)
plot(density(transplant[censored_index,]$time), main = "relapse time density plot for right-censored pa
```

relapse time density plot for right-censored patients

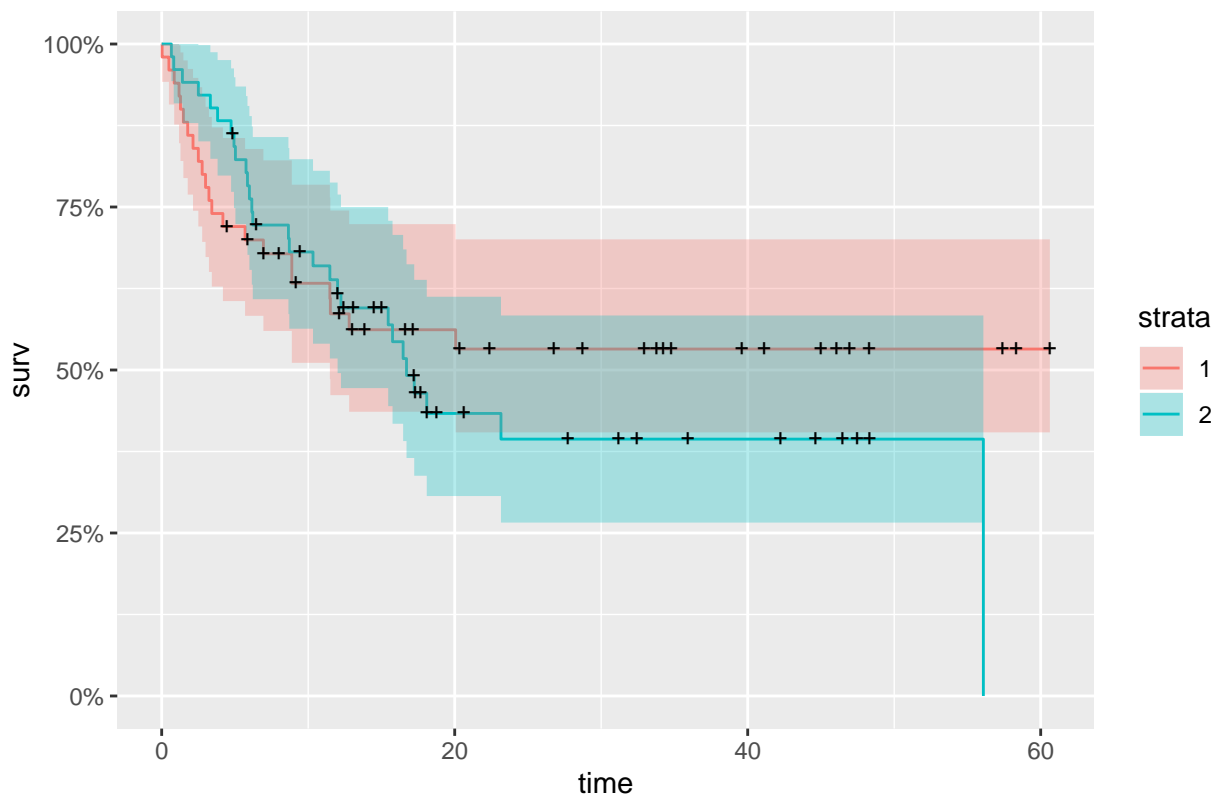


As we can see from the density plot, the right-censored patients, which are alive patients, do not have a fixed censoring time; hence, it is reasonable to assume that the right censoring in this dataset is random.

Problem 2

```
autoplot(survfit(Surv(time, Indicator)~type, data=transplant), conf.int=TRUE, col=c(1,3), main="Kaplan-Meier Survival Plot")
```

Kaplan–Meier estimates for transplant data



We can see from the above plot that there is a minor difference in the transplant survival rate between the two types. In general, Type 1 (allogeneic) transplant seems to be slightly more efficient since Type 2 (autologous) transplant has a large drop in survival after 55 months.

Problem 3

```
model_exp <- survreg(Surv(time, Indicator)~type, data=transplant, dist = "exponential")
summary(model_exp)
```

```
##
## Call:
## survreg(formula = Surv(time, Indicator) ~ type, data = transplant,
##         dist = "exponential")
##               Value Std. Error      z      p
## (Intercept)  3.742      0.213 17.55 <2e-16
## type2       -0.325      0.285 -1.14  0.25
##
## Scale fixed at 1
##
## Exponential distribution
## Loglik(model)= -228   Loglik(intercept only)= -228.6
##  Chisq= 1.31 on 1 degrees of freedom, p= 0.25
## Number of Newton-Raphson Iterations: 5
## n= 101
```

We can see from the output that type 2 (autologous), on average, has 0.325 lower survival than type 1 (allogeneic) transplant, and this difference is not significant as the associated p-value is 0.25. This result agrees

with my intuition from last point since the previous plot only showed type 2 (autologous) transplant has a slightly lower survival than type 1.

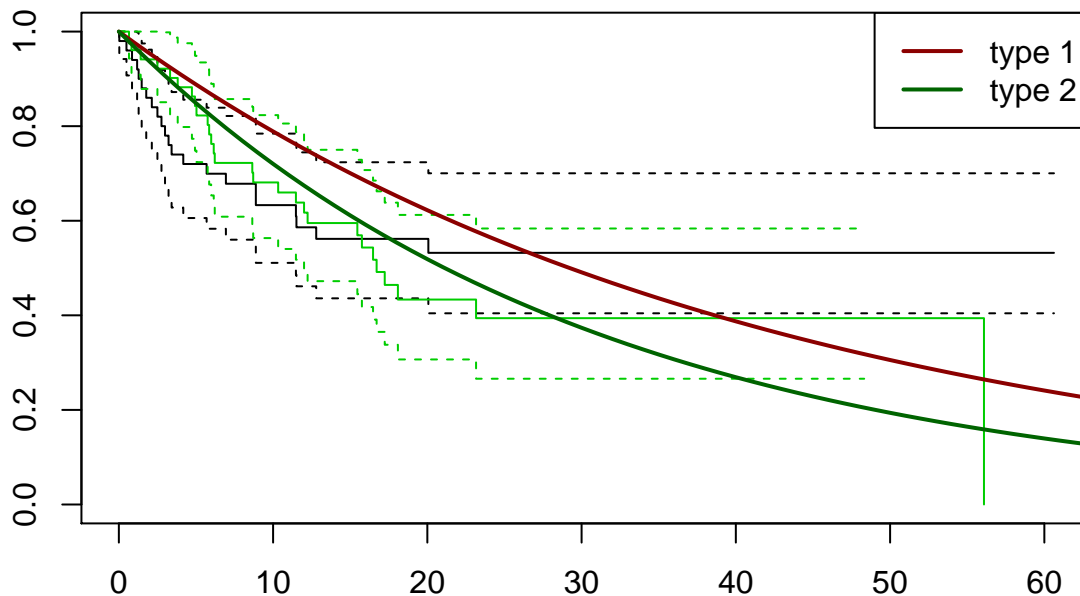
Problem 4

From the summary table, we see that the likelihood ratio statistics gives a p-value of 0.25, indicating that the difference between type 1 and type 2 treatments is not significant. It also agrees with the insignificant coefficient for type2. This conclusion depends on the exponential model assumption.

Problem 5

```
plot(survfit(Surv(time, Indicator)~type, data=transplant), conf.int=TRUE, col=c(1,3), main="Exponential
x <- seq(0, 70, 1)
lines(x, 1-pexp(x, exp(-model_exp$coefficients[1])), col="darkred", lwd=2)
lines(x, 1-pexp(x, exp(-sum(model_exp$coefficients))), col="darkgreen", lwd=2)
legend("topright", legend = c("type 1", "type 2"),
      col=c("darkred", "darkgreen"), lwd = 2)
```

Exponential v.s. K-M fits



We observe that the exponential model does not fit very well. For both transplant type, we see some departures from the pointwise confidence intervals of the K-M estimates. Although the exponential model fits type 2 better, the fitted line of type 2 deviates from the confidence intervals at around time = 28.

Problem 6

```
model_wei <- survreg(Surv(time, Indicator)~type,data=transplant)
summary(model_wei)

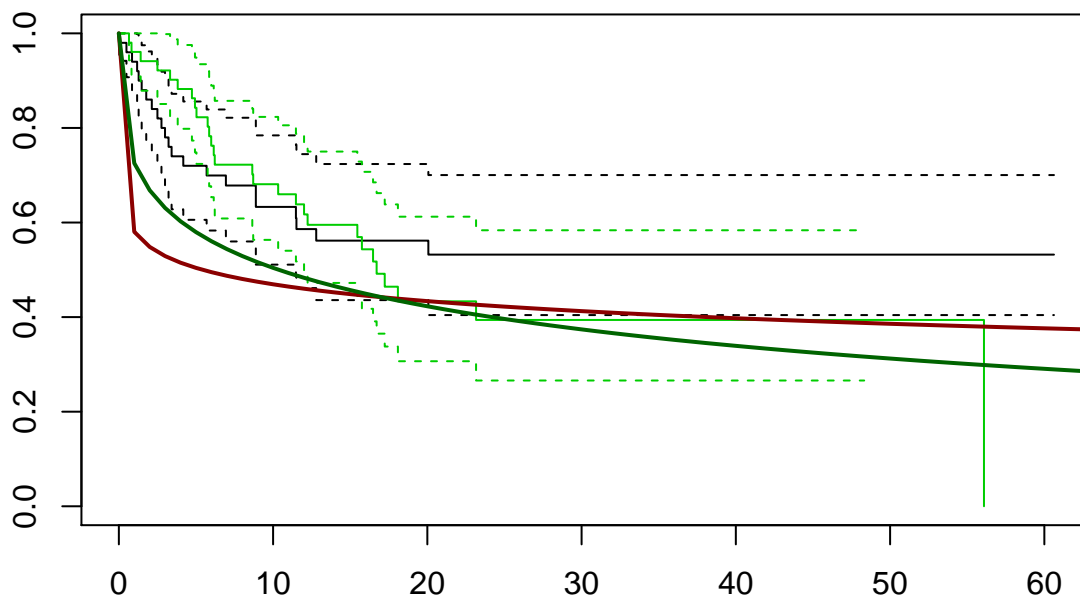
##
## Call:
## survreg(formula = Surv(time, Indicator) ~ type, data = transplant)
```

```
##              Value Std. Error      z      p
## (Intercept)  3.968      0.330 12.02 <2e-16
## type2       -0.374      0.420 -0.89 0.3742
## Log(scale)   0.388      0.123  3.15 0.0016
##
## Scale= 1.47
##
## Weibull distribution
## Loglik(model)= -222   Loglik(intercept only)= -222.4
##  Chisq= 0.8 on 1 degrees of freedom, p= 0.37
## Number of Newton-Raphson Iterations: 5
## n= 101

type1_index <- which(transplant$type==1)
fit.wei1<-survreg(Surv(time, Indicator)~type, data=transplant[type1_index,])
fit.wei2<-survreg(Surv(time, Indicator)~type, data=transplant[-type1_index,])
gamma1=1/exp(fit.wei1$scale)
gamma2=1/exp(fit.wei2$scale)

plot(survfit(Surv(time, Indicator)~type, data=transplant), conf.int=TRUE, col=c(1,3), main="Weibull (Sp
x <- seq(0, 70, 1)
lines(x, 1-pweibull(x, gamma1, exp(coef(fit.wei1)[1])), col="darkred",lwd=2)
lines(x, 1-pweibull(x, gamma2, exp(coef(fit.wei2)[1])), col="darkgreen",lwd=2)
```

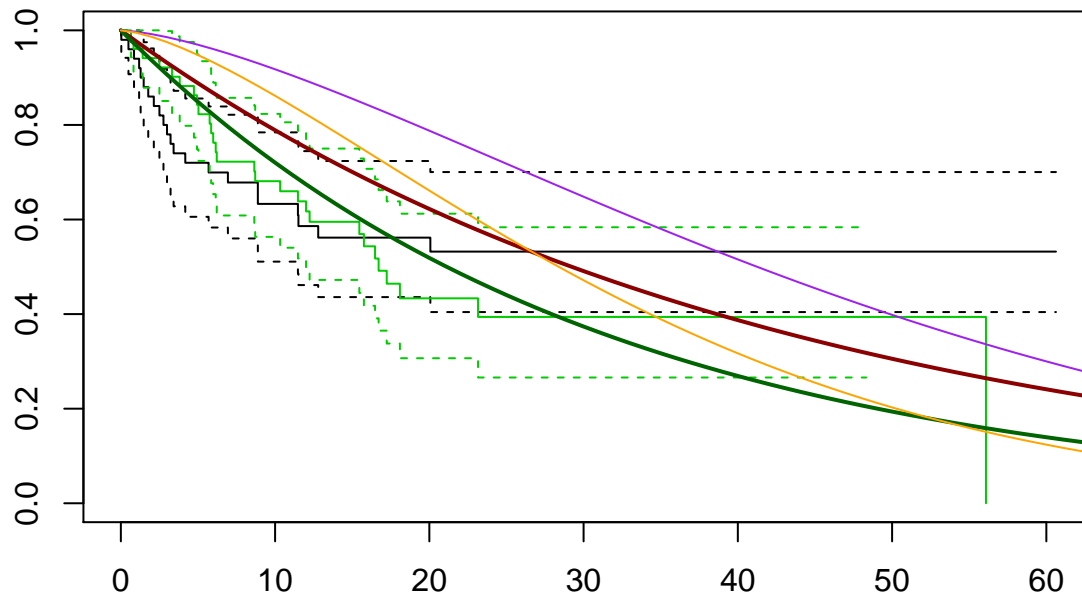
Weibull (Split data) v.s. K-M fits



When we split the data according to the transplant types as what we did in the lecture notes, we can see the model fits even worse; hence, we may want to use the whole dataset.

```
plot(survfit(Surv(time, Indicator)~type, data=transplant), conf.int=TRUE, col=c(1,3), main="Weibull (wh
x <- seq(from=0,to=70,by=1)
lines(x, 1-pexp(x, exp(-model_exp$coefficients[1])), col="darkred", lwd=2)
lines(x, 1-pexp(x, exp(-sum(model_exp$coefficients))), col="darkgreen", lwd=2)
lines(x, 1-pweibull(x, model_wei$scale, exp(coef(model_wei)[1])), col="purple")
lines(x, 1-pweibull(x, model_wei$scale, exp(sum(coef(model_wei)))), col="orange")
```

Weibull (whole data) v.s. K-M fits



Still, using the whole dataset does not give a better fit. This result agrees with what we have from the weibull fit model as we can see the p-value of the model is 0.37, indicating that the weibull does not fit the model well and there is no improvement from the exponential model.