# STAT 5703 Homework 2 Exercise 1

Shijie He(sh3975), Yunjun Xia(yx2569), Shuyu Huang(sh3967)

3/1/2020

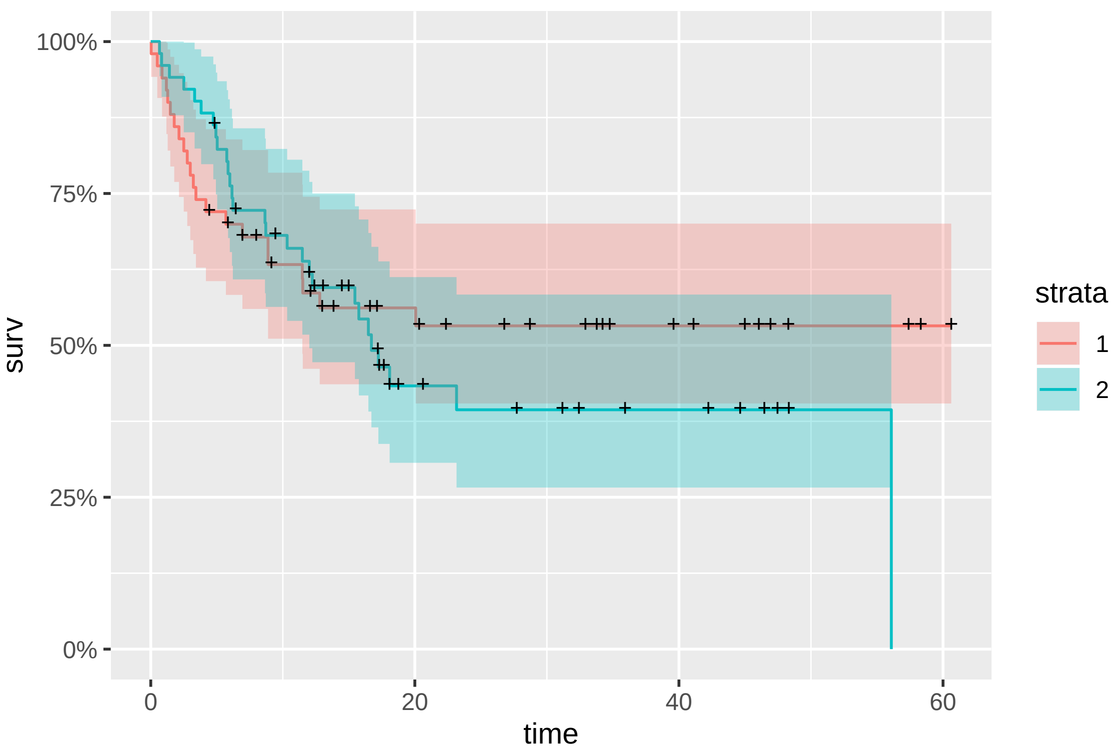
## Part 1

```
transplant <- read_table2("transplant.txt", col_names = FALSE)[-(1:7),(1:3)]
colnames(transplant) = c("t", "type", "survive")
transplant = as.data.frame(transplant)
transplant %<>% mutate_if(is.character, as.numeric)
```

The right censoring in the data can be assumed to be random since the right censored data have different censoring time.

## Part 2

```
autoplot(survfit(Surv(t,survive)~type,data=transplant))
```




There is a significant gap between allogeneic and autologous transplant, so there is difference between the groups. Type I transplant has a higher survival time, thus, it is more efficient.

## Part 3

```r
mod = survreg(Surv(t, survive) ~ type, data=transplant, dist = "exponential")
summary(mod)
```

```
##
## Call:
## survreg(formula = Surv(t, survive) ~ type, data = transplant,
##     dist = "exponential")
##              Value Std. Error     z      p
## (Intercept)  4.066      0.466  8.72 <2e-16
## type        -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
```

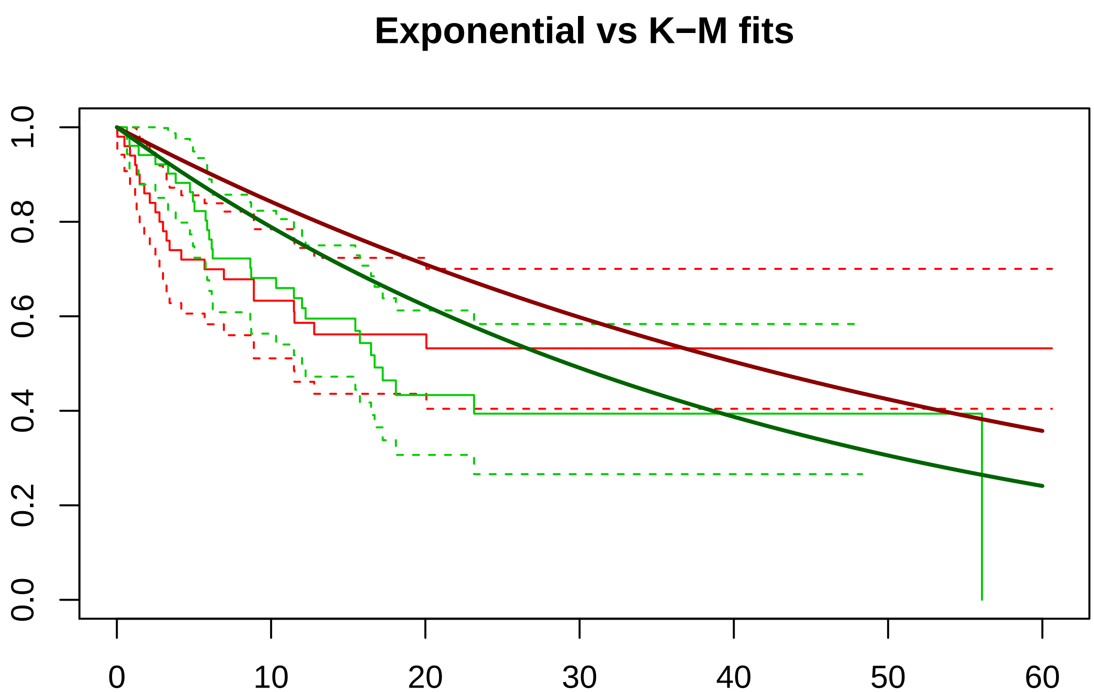
Type II has the estimated coefficient of -0.325. This fitted parameter is negative, which agrees with that type II has a lower survival rate.

## Part 4

The likelihood-ratio chi-square is 1.31, which means that there is no significant difference between the two groups. We can also see this from the p-value: 0.25 does not indicates a significant difference. This conclusion depends on parametric model assumption since we assume the model to be exponential model.

## Part 5

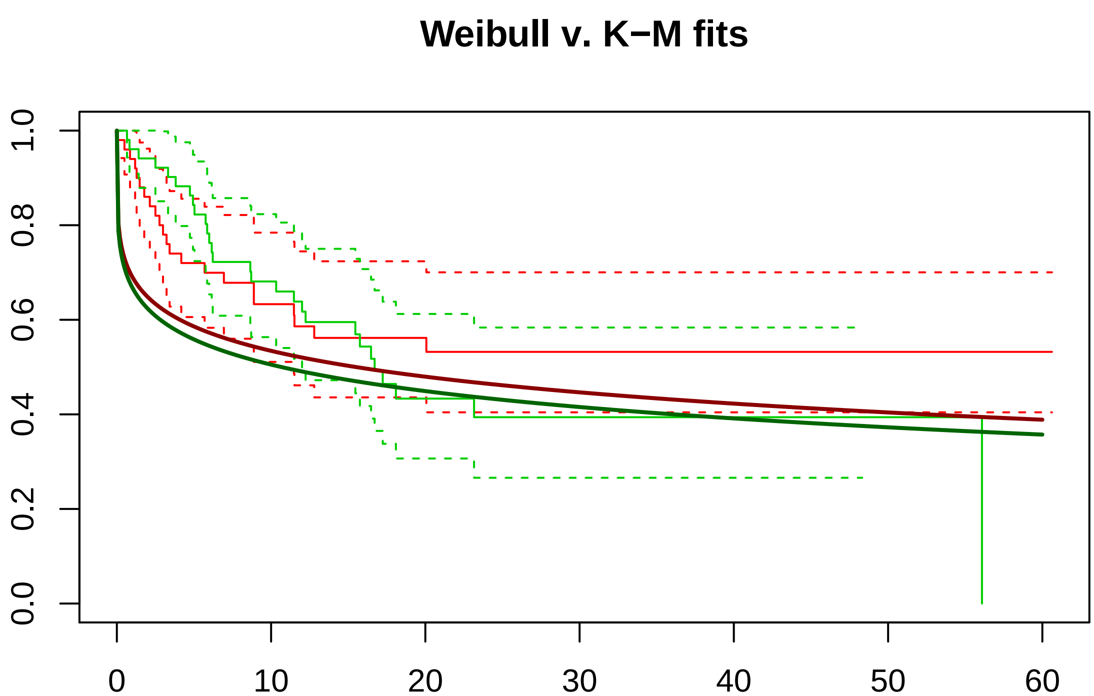
```r
plot(survfit(Surv(t,survive) ~ type, data = transplant), conf.int=TRUE, col=c(2,3), main = "Exponential
x <- seq(from=0, to=60, by=0.1)

lines(x, 1-pexp(x,exp(-coef(mod)[1])), col="darkred", lwd=2)
lines(x, 1-pexp(x,exp(-sum(coef(mod)))), col="darkgreen", lwd=2)
```




The exponential lines give an overall shape of the data, but it does not fit the data well. We can see that the line goes beyond the confidence interval at the right tail.

### Part 6

```
mod_wei <- survreg(Surv(t,survive) ~ type, data = transplant)
gamma = 1/exp(mod_wei$scale)
plot(survfit(Surv(t,survive) ~ type, data = transplant), conf.int=TRUE, col=c(2,3), main="Weibull v. K-
lines(x,1-pweibull(x, gamma, exp(coef(mod_wei)[1])),col="darkred",lwd=2)
lines(x,1-pweibull(x, gamma, exp(sum(coef(mod_wei)))),col="darkgreen",lwd=2)
```



The model fits the data better. The lines become flatter as time increases, and the shape of the curves fits the data better. As we know, exponential distribution is a special case of Weibull distribution. Thus, a more general model will definitely gives a better fit of the data.

```r
summary(mod_wei)
```

```
##
## Call:
## survreg(formula = Surv(t, survive) ~ type, data = transplant)
##              Value Std. Error     z      p
## (Intercept)  4.342      0.698  6.22  5e-10
## type        -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
```

The scale value here is closed to 1, which makes it an exponential distribution. Hence, we could conclude that there is no signficant difference between Weibull distribution and exponential distribution fit. We could further perform a two sample test to test if there is a significance improvement after using Weibull distribution, but for now, the value of scale of 1.47 gives enough evidence to conclude the little improvement.