

d. Plot the two survival curves and comment the results.

First, I construct the data. For treatment, 0 is Treatment A, and 1 is Treatment B. For status, 1 is fail, and 0 is censored. Since the DFS for S7 in Treatment B is >36, I just set it to be 37.

Below is the plot of the two survival curves.

```
library(survival)
DFS <- c(1, 3, 3, 4, 4, 6, 6, 6, 22, 7, 9, 8, 9, 9, 10, 11, 15, 25,
        6, 5, 7, 15, 24, 8, 3, 13, 14, 25, 27, 37)
status <- c(rep(1,29),0)
treatment <- c(rep(0,18), rep(1,12))
ova.cancer <- as.data.frame(cbind(DFS, status, treatment))
ova.cancer
```

```
##      DFS status treatment
## 1      1      1          0
## 2      3      1          0
## 3      3      1          0
## 4      4      1          0
## 5      4      1          0
## 6      6      1          0
## 7      6      1          0
## 8      6      1          0
## 9     22      1          0
## 10     7      1          0
## 11     9      1          0
## 12     8      1          0
## 13     9      1          0
## 14     9      1          0
## 15    10      1          0
## 16    11      1          0
## 17    15      1          0
## 18    25      1          0
## 19     6      1          1
## 20     5      1          1
## 21     7      1          1
## 22    15      1          1
## 23    24      1          1
## 24     8      1          1
## 25     3      1          1
## 26    13      1          1
## 27    14      1          1
## 28    25      1          1
## 29    27      1          1
## 30    37      0          1
```

```
attach(ova.cancer)
```

```
## The following objects are masked _by_ .GlobalEnv:
##
##      DFS, status, treatment
```

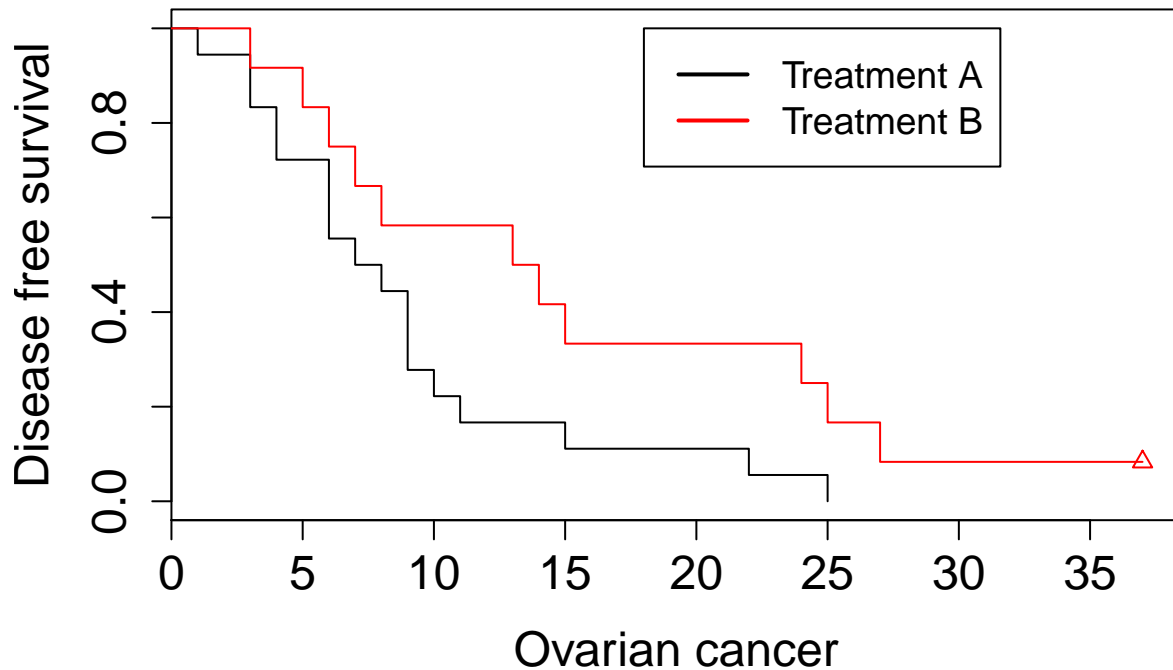
```
fit <- survfit(Surv(DFS, status) ~ treatment, data=ova.cancer)
summary(fit)
```

```
## Call: survfit(formula = Surv(DFS, status) ~ treatment, data = ova.cancer)
```

```
##
##               treatment=0
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    1     18      1  0.9444  0.0540    0.84434    1.000
##    3     17      2  0.8333  0.0878    0.67779    1.000
##    4     15      2  0.7222  0.1056    0.54231    0.962
##    6     13      3  0.5556  0.1171    0.36752    0.840
##    7     10      1  0.5000  0.1179    0.31502    0.794
##    8      9      1  0.4444  0.1171    0.26516    0.745
##    9      8      3  0.2778  0.1056    0.13188    0.585
##   10      5      1  0.2222  0.0980    0.09364    0.527
##   11      4      1  0.1667  0.0878    0.05932    0.468
##   15      3      1  0.1111  0.0741    0.03008    0.410
##   22      2      1  0.0556  0.0540    0.00827    0.373
##   25      1      1  0.0000    NaN          NA          NA
##
##               treatment=1
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    3     12      1  0.9167  0.0798    0.7729    1.000
##    5     11      1  0.8333  0.1076    0.6470    1.000
##    6     10      1  0.7500  0.1250    0.5410    1.000
##    7      9      1  0.6667  0.1361    0.4468    0.995
##    8      8      1  0.5833  0.1423    0.3616    0.941
##   13      7      1  0.5000  0.1443    0.2840    0.880
##   14      6      1  0.4167  0.1423    0.2133    0.814
##   15      5      1  0.3333  0.1361    0.1498    0.742
##   24      4      1  0.2500  0.1250    0.0938    0.666
##   25      3      1  0.1667  0.1076    0.0470    0.591
##   27      2      1  0.0833  0.0798    0.0128    0.544
```

```
plot(fit, firstx=0, mark.time=T, mark=c(1,2), col=c(1,2), ylim=c(0,1), xlab="Ovarian cancer", ylab="Dis
legend(x=18,y=1,legend=c("Treatment A","Treatment B"), col=c(1,2),lwd=2,cex=1.2)
```

Kaplen Meier Curves



From the plot, the survival distribution for women with Treatment A and Treatment B are not so much different.

I also use the Log-Rank Test to prove this result. $H_0: S_1(t)=S_2(t)$; $H_1: S_1(t) = [S_2(t)]^\theta$.

```
survdifff(Surv(DFS,status) ~ treatment, data=ova.cancer)
```

```
## Call:
## survdifff(formula = Surv(DFS, status) ~ treatment, data = ova.cancer)
##
##               N Observed Expected (O-E)^2/E (O-E)^2/V
## treatment=0 18         18      13.2       1.73      3.8
## treatment=1 12         11      15.8       1.45      3.8
##
## Chisq= 3.8  on 1 degrees of freedom, p= 0.0512
```

From the output, since the p-value is $0.0512 > 0.05$, we can not reject the H_0 . We conclude the two survival curves are not statistically different, which improves the plot we draw.

I also plot the log comparison.

```
plot(log(fit$time),log(-log(fit$surv)))
lines(log(fit$time[1:12]),log(-log(fit$surv[1:12])))
lines(log(fit$time[13:24]),log(-log(fit$surv[13:24])),col=2)
legend(x=0.1,y=1,legend=c("1:12 Index","13:24 Index"), col=c(1,2),lwd=2,cex=1.2)
```

