

# IS624 Week 7

*Aaron Palumbo*

*Monday, July 20, 2015*

## Contents

Question 1: Plotting/Comparing Survival Curves Part I . . . . .	1
Question 2: Plotting/Comparing Survival Curves Part . . . . .	3
Question 3: Developing a Cox Model . . . . .	4
Univariate . . . . .	4

## Question 1: Plotting/Comparing Survival Curves Part I

Load in the ovarian data-set:

```
library(survival)
data(ovarian)
head(ovarian)
```

```
##   futime fustat      age resid.ds rx ecog.ps
## 1     59      1 72.3315        2  1        1
## 2    115      1 74.4932        2  1        1
## 3    156      1 66.4658        2  1        2
## 4    421      0 53.3644        2  2        1
## 5    431      1 50.3397        2  1        1
## 6    448      0 56.4301        1  1        2
```

This data is from a randomised trial comparing two treatments for ovarian cancer. Look in <http://cran.r-project.org/web/packages/survival/survival.pdf>, the variables are as follows:

**futime:** survival or censoring time  
**fustat:** censoring status  
**age:** in years  
**resid.ds:** residual disease present (1-no, 2-yes)  
**rx:** treatment group  
**ecog.ps:** ECOG performance status (1 is better)

Compare the two treatment groups (**treatment1** versus **treatment2**). Plot the respective survival curves, indicating censored subjects. You can distinguish between the two groups using different colors or different line formats or both. Label both x- and y-axes. Add a suitable title. Also, add a legend indicating which line corresponds to which line format.

Finally compare the two survival curves (log-rank) and add a p-value to the bottom-right of the plot.

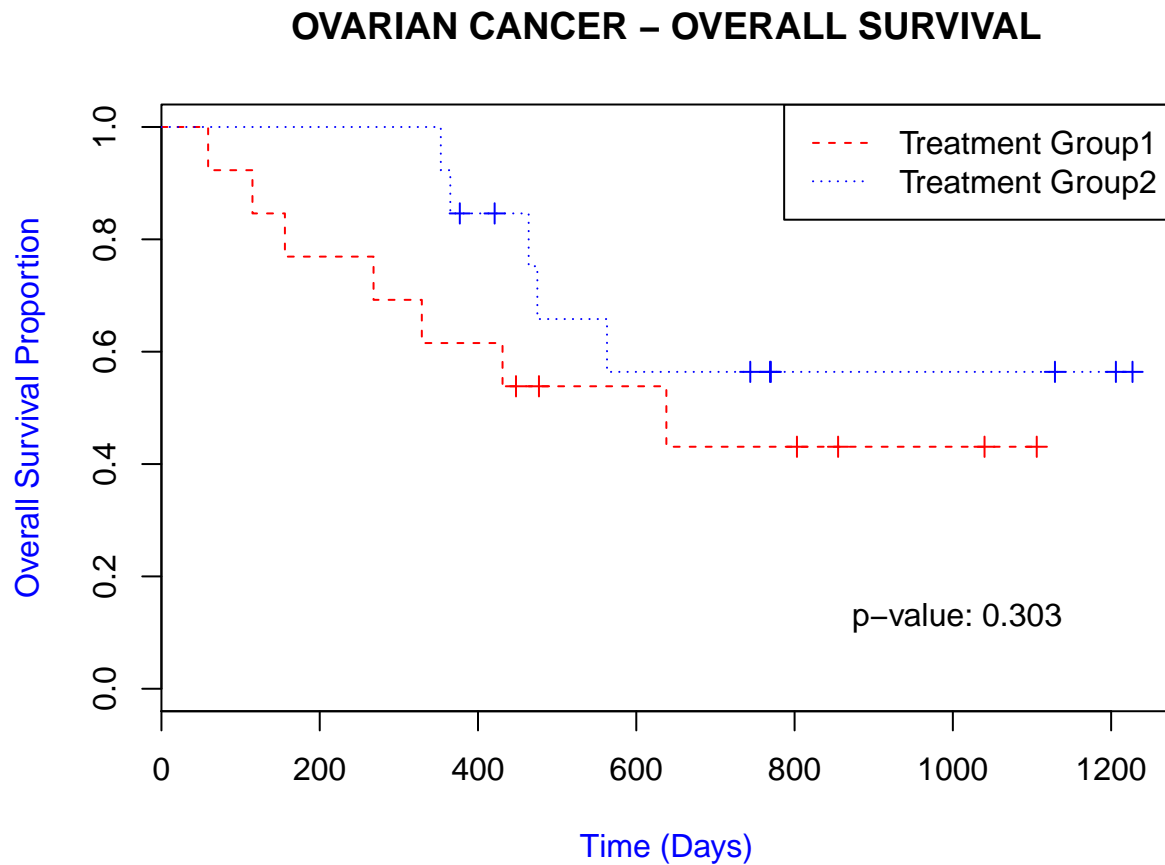
HINT: You add tow or more legends to the plot; to add a legend without a border use box.col="white".

You should achieve something like this:

```
fit <- survfit(Surv(futime, fustat)~rx, data=ovarian)
survdif(Surv(futime,fustat)~rx, data=ovarian)
```

```
## Call:
## survdiff(formula = Surv(futime, fustat) ~ rx, data = ovarian)
##
##      N Observed Expected (O-E)^2/E (O-E)^2/V
## rx=1 13      7      5.23    0.596    1.06
## rx=2 13      5      6.77    0.461    1.06
##
## Chisq= 1.1  on 1 degrees of freedom, p= 0.303
```

```
plot(fit, main="OVARIAN CANCER - OVERALL SURVIVAL",
     xlab="Time (Days)", ylab="Overall Survival Proportion", col.lab='blue',
     col=c("red", "blue"), lty=2:3)
legend("topright", lty=2:3, col=c("red", "blue"),
     legend=c("Treatment Group1",
              "Treatment Group2"))
legend(x=800, y=.2, legend="p-value: 0.303", box.col="white")
```



## Question 2: Plotting/Comparing Survival Curves Part

Load in the Leukaemia-free survival/transplant data-set:

```
library(KMsurv)
data(alloauto)
head(alloauto)
```

```
##      time type delta
## 1 0.030     1      1
## 2 0.493     1      1
## 3 0.855     1      1
## 4 1.184     1      1
## 5 1.283     1      1
## 6 1.480     1      1
```

This data considers two transplant types in relation to leukaemia-free survival. Look in <http://cran.r-project.org/web/packages/KMsurv/KMsurv.pdf>, the variables are as follows:

**Time:** Time to death or relapse, months

**Type:** Type of transplant (1-allogeneic, 2-autologous)

**Delta:** Leukemia-free survival indicator (0=alive without relapse, 1=deceased or relapse)

Compare the two transplant types (**allogeneic** vs **autologous**). Plot the respective relapse-free survival curves, indicating censored subjects. You can distinguish between the two groups using different colors or different line formats or both. Label both x- and y-axes. Add a suitable title. Also, add a legend indicating which line corresponds to which line format.

Finally compare the two survival curves (log-rank) and add a p-value to the bottom-right of the plot.

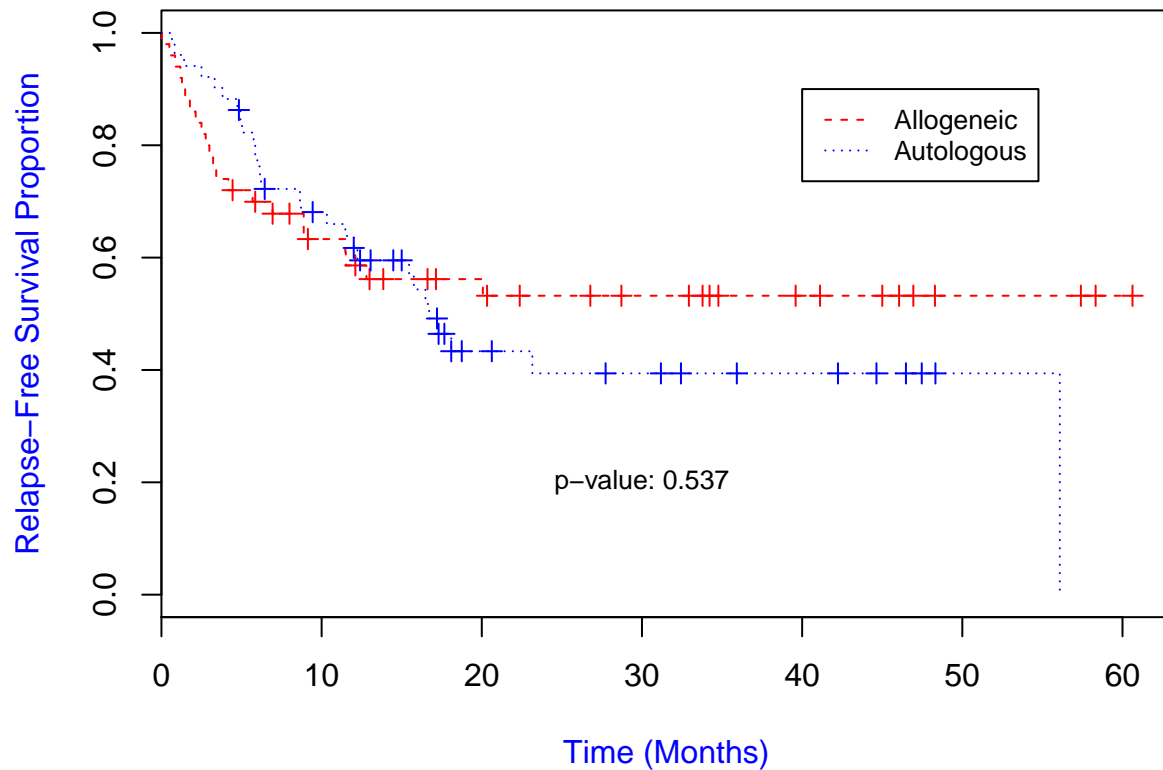
You should achieve something like this:

```
fit <- survfit(Surv(time, delta)~type, data=alloauto)
survdiff(Surv(time, delta)~type, data=alloauto)
```

```
## Call:
## survdiff(formula = Surv(time, delta) ~ type, data = alloauto)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## type=1 50      22      24.2      0.195      0.382
## type=2 51      28      25.8      0.182      0.382
##
##  Chisq= 0.4  on 1 degrees of freedom, p= 0.537
```

```
plot(fit, main="Leukemia - Free Survival",
     xlab="Time (Months)", ylab="Relapse-Free Survival Proportion", col.lab='blue',
     col=c("red", "blue"), lty=2:3)
legend(40, 0.9, lty=2:3, col=c("red", "blue"),
      legend=c("Allogeneic", "Autologous"), cex=0.8)
text(30, .2, labels="p-value: 0.537", cex=0.8)
```

## Leukemia – Free Survival



### Question 3: Developing a Cox Model

Look at the hmoiv data-set again.

Recode the age variables (see below) and repeat the univariate and multivariate analyses.

```
library(car)
hmohiv <- read.table("http://www.ats.ucla.edu/stat/R/examples/asa/hmohiv.csv",
  sep="," , header = TRUE)
```

#### Univariate

```
agecat <- recode(hmohiv$age, "20:29='D'; 30-34='B'; 35:39='C'; 40:54='A'", as.factor=TRUE)
coxph(formula=Surv(time, censor)~agecat, method="efron", data=hmohiv)
```

```
## Call:
## coxph(formula = Surv(time, censor) ~ agecat, data = hmohiv, method = "efron")
##
##
```

```

##          coef exp(coef) se(coef)      z    p
## agecat31  0.1943    1.2144   0.7116   0.27 0.78
## agecat32  0.6273    1.8726   0.6155   1.02 0.31
## agecat33 -0.1535    0.8577   0.6734  -0.23 0.82
## agecat34  0.0922    1.0966   0.6179   0.15 0.88
## agecatA   0.8785    2.4073   0.5467   1.61 0.11
## agecatC   0.2923    1.3395   0.5515   0.53 0.60
## agecatD  -1.0540    0.3486   0.6522  -1.62 0.11
##
## Likelihood ratio test=23.1  on 7 df, p=0.00164
## n= 100, number of events= 80

```