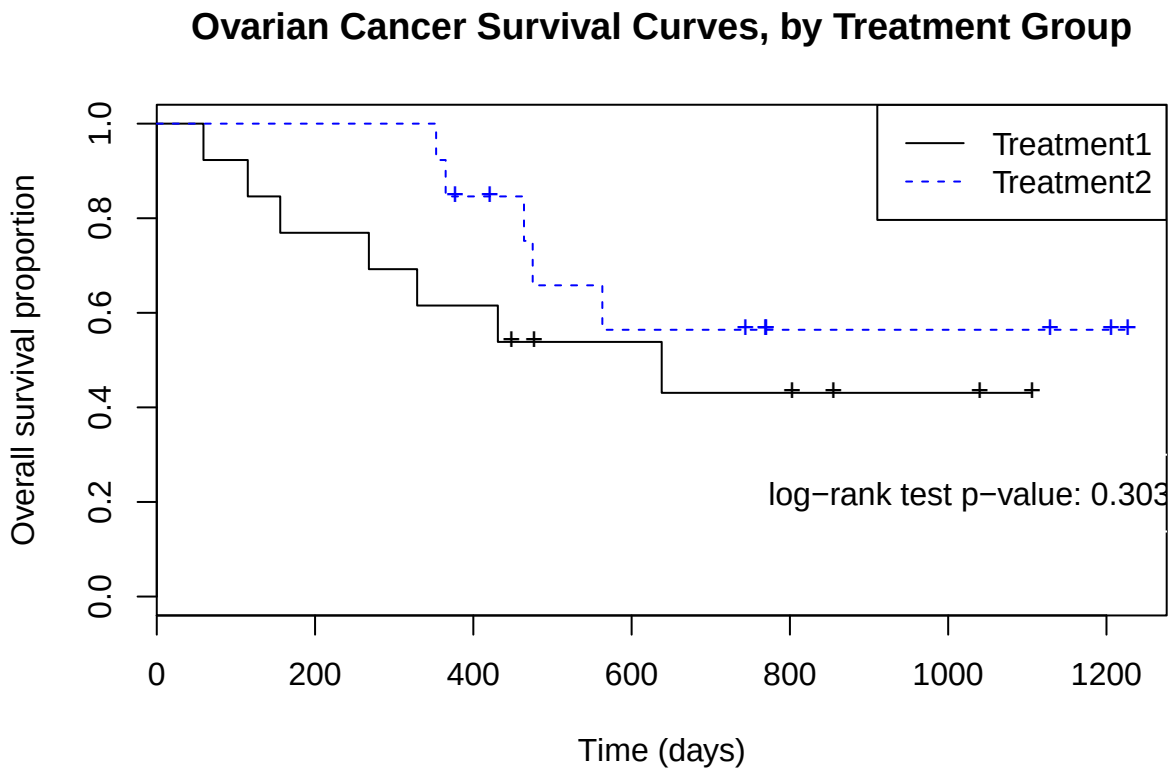
IS624: Homework 7

Brian Chu | July 25, 2015

```
library(survival)
library(KMsurv)
library(MASS)
```

Question 1: Plotting/Comparing Survival Curves Part I

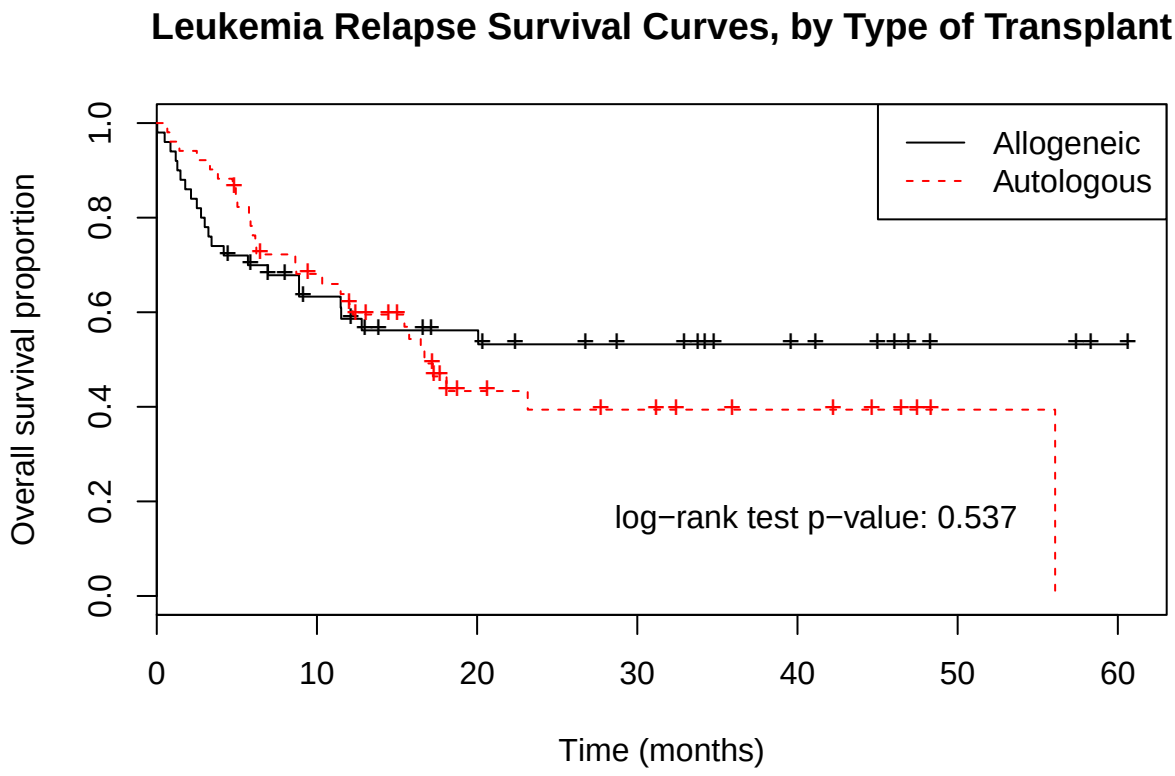
```
data(ovarian)
attach(ovarian)
ovarian_survFit <- survfit(Surv(futime, fustat) ~ rx)
plot(ovarian_survFit, main="Ovarian Cancer Survival Curves, by Treatment Group",
     xlab="Time (days)", ylab="Overall survival proportion",
     col=c("black", "blue"), lty=1:2, mark="+")
legend("topright", c("Treatment1", "Treatment2"), col=c("black", "blue"), lty=1:2)
survFit_diff <- survdiff(Surv(futime, fustat) ~ rx)
legend(700, 0.3, "log-rank test p-value: 0.303", box.col="white")
```




```
detach(ovarian)
```

Question 2: Plotting/Comparing Survival Curves Part II

```
data(alloauto)
attach(alloauto)
alloauto_survFit <- survfit(Surv(time, delta) ~ type)
plot(alloauto_survFit, main="Leukemia Relapse Survival Curves, by Type of Transplant",
     xlab="Time (months)", ylab="Overall survival proportion",
     col=c("black", "red"), lty=1:2, mark="+")
legend("topright", c("Allogeneic", "Autologous"), col=c("black", "red"), lty=1:2)
survFit_diff <- survdiff(Surv(time, delta) ~ type)
legend(25, 0.25, "log-rank test p-value: 0.537", box.col="white")
```




```
detach(alloauto)
```

Question 3: Developing a Cox Model

```
#Univariate analysis
agecat2 <- recode(age, "20:29='D'; 30:34='B'; 35:39='C'; 40:54='A'", as.factor=T)
agecat2.coxph <- coxph(Surv(time, censor)~agecat2, method="efron")
# summary(drug.coxph)
summary(agecat2.coxph)
```

```
## Call:
```

```
## coxph(formula = Surv(time, censor) ~ agecat2, method = "efron")
##
##   n= 100, number of events= 80
##
##             coef exp(coef) se(coef)      z Pr(>|z|)
## agecat2B -0.7114    0.4909   0.2870 -2.479   0.0132 *
## agecat2C -0.5807    0.5595   0.3122 -1.860   0.0629 .
## agecat2D -1.9144    0.1474   0.4679 -4.091 4.29e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##          exp(coef) exp(-coef) lower .95 upper .95
## agecat2B    0.4909      2.037   0.27971    0.8617
## agecat2C    0.5595      1.787   0.30344    1.0316
## agecat2D    0.1474      6.783   0.05892    0.3689
##
## Concordance= 0.642  (se = 0.04 )
## Rsquare= 0.189   (max possible= 0.997 )
## Likelihood ratio test= 20.92  on 3 df,   p=0.0001091
## Wald test            = 17.85  on 3 df,   p=0.0004724
## Score (logrank) test = 19.83  on 3 df,   p=0.0001843
```

```
#Multivariate analysis
agecat2_drug.coxph <- coxph( Surv(time, censor) ~ agecat2+drug, method="efron")
summary(agecat2_drug.coxph)
```

```
## Call:
## coxph(formula = Surv(time, censor) ~ agecat2 + drug, method = "efron")
##
##   n= 100, number of events= 80
##
##             coef exp(coef) se(coef)      z Pr(>|z|)
## agecat2B -0.7582    0.4685   0.2899 -2.615  0.00891 **
## agecat2C -0.6958    0.4987   0.3151 -2.208  0.02722 *
## agecat2D -2.0152    0.1333   0.4804 -4.195 2.73e-05 ***
## drug      0.8926    2.4415   0.2530  3.527  0.00042 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##          exp(coef) exp(-coef) lower .95 upper .95
## agecat2B    0.4685     2.1345   0.26541    0.8269
## agecat2C    0.4987     2.0053   0.26892    0.9247
## agecat2D    0.1333     7.5019   0.05199    0.3418
## drug        2.4415     0.4096   1.48685    4.0092
##
## Concordance= 0.681  (se = 0.042 )
## Rsquare= 0.284   (max possible= 0.997 )
## Likelihood ratio test= 33.35  on 4 df,   p=1.015e-06
## Wald test            = 28.7  on 4 df,   p=9.012e-06
## Score (logrank) test = 31.57  on 4 df,   p=2.339e-06
```

```
detach(hmohiv)
```

Does using a larger reference group for the age help in any way? What do you notice about the hazard ratios and confidence intervals? Are the p-values or the concordance index affected?

```
uis <-read.table("http://www.ats.ucla.edu/stat/R/examples/asa/uis.csv", sep=",", header = TRUE)
uis_small<-uis[,c(1,2,4,6,8,9,11,12)]
tiny_uis <- uis_small[apply(uis_small,1,function(x)!any(is.na(x))),]
attach(tiny_uis)

#Age univariate
age.coxph <- coxph(Surv(time,censor)~age+strata(site), method="efron", data=tiny_uis)
summary(age.coxph)
```

Question 4: Variable Selection

```
## Call:
## coxph(formula = Surv(time, censor) ~ age + strata(site), data = tiny_uis,
##     method = "efron")
##
##   n= 593, number of events= 481
##
##          coef exp(coef)  se(coef)      z Pr(>|z|)
## age -0.014839  0.985270  0.007393 -2.007   0.0447 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##     exp(coef) exp(-coef) lower .95 upper .95
## age    0.9853      1.015    0.9711    0.9997
##
## Concordance= 0.532  (se = 0.018 )
## Rsquare= 0.007   (max possible= 1 )
## Likelihood ratio test= 4.09  on 1 df,   p=0.04323
## Wald test            = 4.03  on 1 df,   p=0.04473
## Score (logrank) test = 4.03  on 1 df,   p=0.04466
```

```
#Treatment univariate
treat.coxph <- coxph(Surv(time,censor)~treat+strata(site), method="efron", data=tiny_uis)
summary(treat.coxph)
```

```
## Call:
## coxph(formula = Surv(time, censor) ~ treat + strata(site), data = tiny_uis,
##     method = "efron")
##
##   n= 593, number of events= 481
##
##           coef exp(coef) se(coef)    z Pr(>|z|)
## treat -0.23815   0.78808  0.09159 -2.6  0.00932 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
##       exp(coef) exp(-coef) lower .95 upper .95
## treat    0.7881      1.269    0.6586    0.9431
##
## Concordance= 0.543  (se = 0.016 )
## Rsquare= 0.011   (max possible= 1 )
## Likelihood ratio test= 6.76  on 1 df,   p=0.009304
## Wald test            = 6.76  on 1 df,   p=0.00932
## Score (logrank) test = 6.79  on 1 df,   p=0.009158
```

```
#Number of drug treatments univariate
ndrugtx.coxph <- coxph(Surv(time,censor)~ndrugtx+strata(site), method="efron", data=tiny_uis)
summary(ndrugtx.coxph)
```

```
## Call:
## coxph(formula = Surv(time, censor) ~ ndrugtx + strata(site),
##     data = tiny_uis, method = "efron")
##
##   n= 593, number of events= 481
##
##             coef exp(coef) se(coef)     z Pr(>|z|)
## ndrugtx 0.030563  1.031035 0.007581 4.031 5.54e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##         exp(coef) exp(-coef) lower .95 upper .95
## ndrugtx     1.031     0.9699     1.016     1.046
##
## Concordance= 0.549  (se = 0.018 )
## Rsquare= 0.023   (max possible= 1 )
## Likelihood ratio test= 14.06  on 1 df,   p=0.0001773
## Wald test            = 16.25  on 1 df,   p=5.545e-05
## Score (logrank) test = 16.39  on 1 df,   p=5.158e-05
```

```
#Heroin/cocaine univariate
hercoc.coxph <- coxph(Surv(time,censor)~hercoc+strata(site), method="efron", data=tiny_uis)
summary(hercoc.coxph)
```

```
## Call:
## coxph(formula = Surv(time, censor) ~ hercoc + strata(site), data = tiny_uis,
##     method = "efron")
##
##   n= 593, number of events= 481
##
##             coef exp(coef) se(coef)      z Pr(>|z|)
## hercoc -0.07069   0.93175  0.04291 -1.647   0.0995 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##        exp(coef) exp(-coef) lower .95 upper .95
## hercoc    0.9318      1.073    0.8566     1.014
##
## Concordance= 0.523  (se = 0.018 )
```

```
## Rsquare= 0.005   (max possible= 1 )
## Likelihood ratio test= 2.69  on 1 df,   p=0.1009
## Wald test            = 2.71  on 1 df,   p=0.09948
## Score (logrank) test = 2.72  on 1 df,   p=0.09927
```

```
#Multivariate analysis
all.coxph <- coxph(Surv(time,censor)~age+treat+ndrugtx+hercoc+strata(site), method="efron", data=tiny_u
summary(all.coxph)
```

```
## Call:
## coxph(formula = Surv(time, censor) ~ age + treat + ndrugtx +
##     hercoc + strata(site), data = tiny_uis, method = "efron")
##
##   n= 593, number of events= 481
##
##              coef exp(coef)  se(coef)      z Pr(>|z|)
## age     -0.025013  0.975298  0.007733 -3.235  0.00122 **
## treat   -0.243350  0.783997  0.092259 -2.638  0.00835 **
## ndrugtx  0.035692  1.036337  0.007833  4.557 5.19e-06 ***
## hercoc  -0.053960  0.947470  0.044475 -1.213  0.22502
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##         exp(coef) exp(-coef) lower .95 upper .95
## age        0.9753     1.0253    0.9606    0.9902
## treat      0.7840     1.2755    0.6543    0.9394
## ndrugtx    1.0363     0.9649    1.0205    1.0524
## hercoc     0.9475     1.0554    0.8684    1.0338
##
## Concordance= 0.585  (se = 0.018 )
## Rsquare= 0.053   (max possible= 1 )
## Likelihood ratio test= 32.26  on 4 df,   p=1.695e-06
## Wald test            = 34.78  on 4 df,   p=5.15e-07
## Score (logrank) test = 34.9  on 4 df,   p=4.859e-07
```

Make a note of the concordance index for each. Create a multivariate model using all four variables. What is the concordance index?

age: 0.532 treat: 0.543 ndrugtx: 0.549 hercoc: 0.523 all: 0.585

Now run the following (AIC) to determine what variables should be retained:

```
stepAIC(all.coxph)
```

```
## Start:  AIC=4942.54
## Surv(time, censor) ~ age + treat + ndrugtx + hercoc + strata(site)
##
##           Df    AIC
## - hercoc   1 4942.0
## <none>       4942.5
## - treat    1 4947.5
## - age      1 4951.2
## - ndrugtx  1 4958.2
##
```

```
## Step:  AIC=4942.01
## Surv(time, censor) ~ age + treat + ndrugtx + strata(site)
##
##           Df    AIC
## <none>       4942.0
## - treat    1 4947.7
## - age      1 4949.7
## - ndrugtx  1 4959.5

## Call:
## coxph(formula = Surv(time, censor) ~ age + treat + ndrugtx +
##     strata(site), data = tiny_uis, method = "efron")
##
##
##              coef exp(coef) se(coef)     z       p
## age     -0.02363   0.97665  0.00765 -3.09  0.0020
## treat   -0.25453   0.77528  0.09183 -2.77  0.0056
## ndrugtx  0.03703   1.03772  0.00770  4.81 1.5e-06
##
## Likelihood ratio test=30.8  on 3 df, p=9.4e-07
## n= 593, number of events= 481
```

```
summary(stepAIC(all.coxph))
```

```
## Start:  AIC=4942.54
## Surv(time, censor) ~ age + treat + ndrugtx + hercoc + strata(site)
##
##           Df    AIC
## - hercoc   1 4942.0
## <none>       4942.5
## - treat    1 4947.5
## - age      1 4951.2
## - ndrugtx  1 4958.2
##
## Step:  AIC=4942.01
## Surv(time, censor) ~ age + treat + ndrugtx + strata(site)
##
##           Df    AIC
## <none>       4942.0
## - treat    1 4947.7
## - age      1 4949.7
## - ndrugtx  1 4959.5

## Call:
## coxph(formula = Surv(time, censor) ~ age + treat + ndrugtx +
##     strata(site), data = tiny_uis, method = "efron")
##
##   n= 593, number of events= 481
##
##              coef exp(coef)  se(coef)      z Pr(>|z|)
## age     -0.023626  0.976651  0.007655 -3.086  0.00203 **
## treat   -0.254530  0.775280  0.091830 -2.772  0.00558 **
## ndrugtx  0.037027  1.037721  0.007700  4.809 1.52e-06 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##         exp(coef) exp(-coef) lower .95 upper .95
## age        0.9767     1.0239    0.9621    0.9914
## treat      0.7753     1.2899    0.6476    0.9282
## ndrugtx    1.0377     0.9636    1.0222    1.0535
##
## Concordance= 0.587  (se = 0.018 )
## Rsquare= 0.051   (max possible= 1 )
## Likelihood ratio test= 30.79  on 3 df,   p=9.405e-07
## Wald test            = 33.49  on 3 df,   p=2.537e-07
## Score (logrank) test = 33.65  on 3 df,   p=2.346e-07
```

What variables are retained? What is the concordance index of the final model?

age, treat, ndrugtx 0.587