IS624 - Assignment7

James Quacinella

07/21/2015

Contents

Question 1

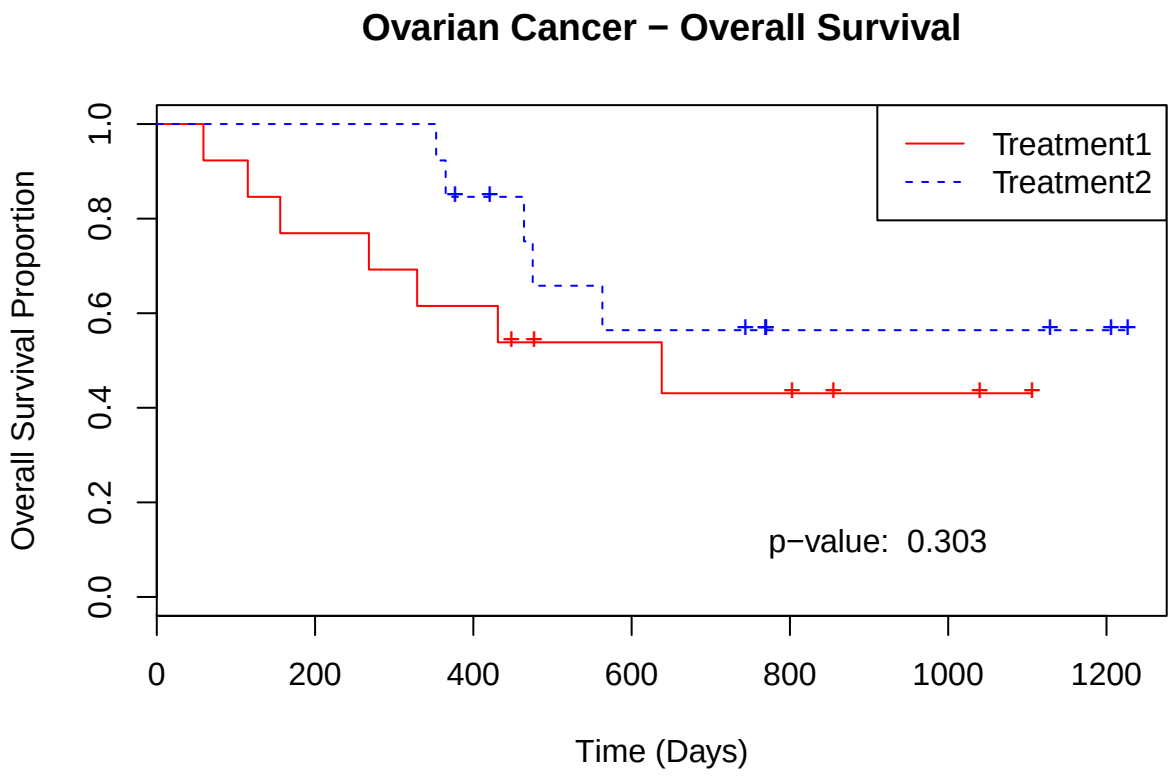
Load in the ovarian data-set, which is data from a randomised trial comparing two treatments for ovarian cancer. Compare the two treatment groups (treatment 1 versus treatment 2). Plot the respective survival curves, indicating censored subjects. You can distinguish between the two groups using different colours 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.

Answer:

Following the code given in the reading:

```
# Load data and create model
attach(ovarian)
ovarian.survFit <- survfit(Surv(futime, fustat) ~ rx)

# Plot survival curves
plot(ovarian.survFit, main="Ovarian Cancer - Overall Survival", xlab="Time (Days)", ylab="Overall Survi
legend("topright", c("Treatment1", "Treatment2"), col=c("red", "blue"), lty=1:2)
legend(700, 0.2, paste("p-value: ", 0.303), box.col="white")
```




```
# Log-rank test p-value (commenting out for formatting)
#survdiff(Surv(futime, fustat) ~ rx) # WTH? No way to get p value??
detach(ovarian)
```

Question 2

This data considers two transplant types in relation to leukaemia-free survival.

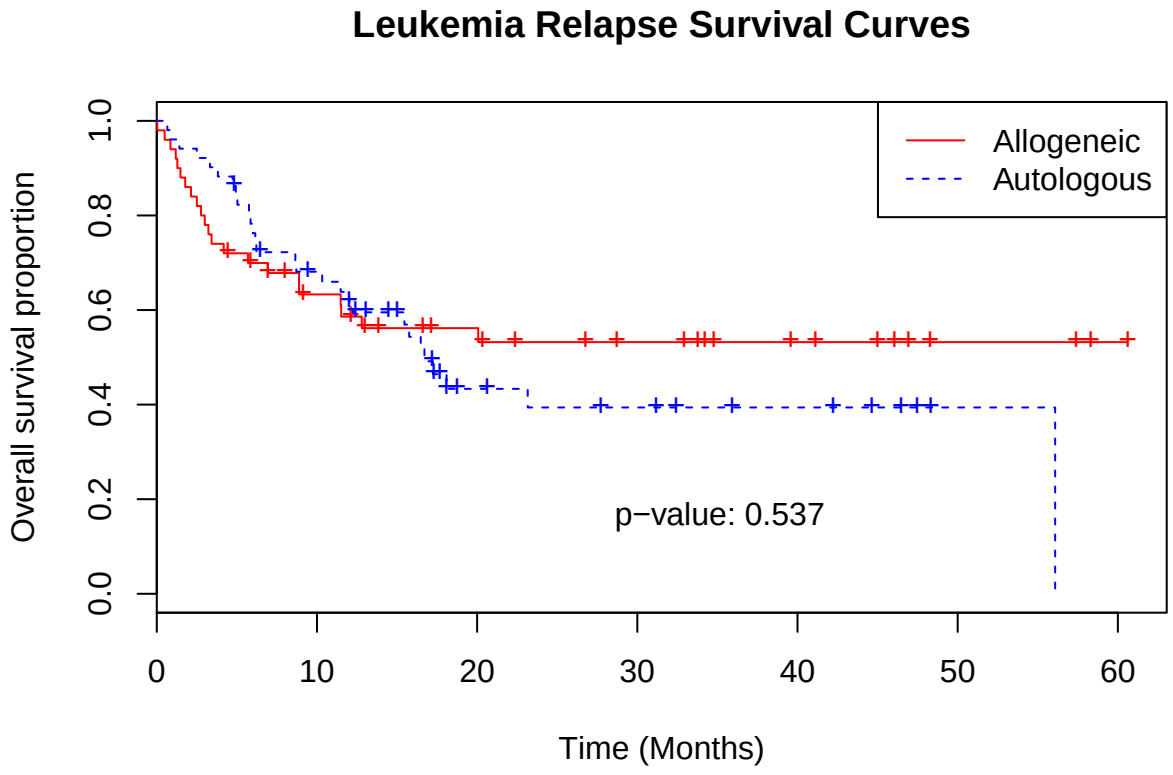
Compare the two transplant types (allogeneic vs autologous). Plot the respective relapse-free survival curves, indicating censored subjects. You can distinguish between the two groups using different colours or different line formats or both. Label both x- and y-axes. Add a suitable title. Also, add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

Answer:

Same methodology as above:

```
# Load data and create model
data(alloauto)
attach(alloauto)
alloauto_survFit <- survfit(Surv(time, delta) ~ type)

# Plot survival curves
plot(alloauto_survFit, main="Leukemia Relapse Survival Curves", xlab="Time (Months)", ylab="Overall sur
legend("topright", c("Allogeneic", "Autologous"), col=c("red", "blue"), lty=1:2)
legend(25, 0.25, "p-value: 0.537", box.col="white")
```




```
# Log-rank test p-value (commenting out for formatting)
#survdiff(Surv(time, delta) ~ type)
detach(alloauto)
```

Question 3

Look at the hmoiv data-set again. Recode the age variables (see below) and repeat the univariate and multivariate analyses.

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

# I think the book has an error in the order of letters but it doesnt really matter
agecat <- recode(age, "20:29='A'; 30:34='B'; 35:39='C';40:54='D'", as.factor=T)
agecat.new <- recode(age, "20:29='D'; 30:34='B'; 35:39='C';40:54='A'", as.factor=T)
```

```
summary(coxph(formula = Surv(time, censor) ~ agecat, method = "efron"))
print("---")
summary(coxph(formula = Surv(time, censor) ~ agecat.new, method = "efron"))
```

```
## Call:
## coxph(formula = Surv(time, censor) ~ agecat, method = "efron")
##
##   n= 100, number of events= 80
##
##           coef exp(coef) se(coef)     z Pr(>|z|)
## agecatB 1.2030    3.3301   0.4503 2.672  0.00755 **
## agecatC 1.3337    3.7951   0.4580 2.912  0.00359 **
## agecatD 1.9144    6.7831   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
## agecatB     3.330     0.3003     1.378     8.049
## agecatC     3.795     0.2635     1.547     9.313
## agecatD     6.783     0.1474     2.711    16.971
##
## 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
##
## [1] "---"
## Call:
## coxph(formula = Surv(time, censor) ~ agecat.new, method = "efron")
##
##   n= 100, number of events= 80
##
##                coef exp(coef) se(coef)      z Pr(>|z|)
## agecat.newB -0.7114    0.4909   0.2870 -2.479   0.0132 *
## agecat.newC -0.5807    0.5595   0.3122 -1.860   0.0629 .
## agecat.newD -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
```

```
## agecat.newB    0.4909      2.037   0.27971    0.8617
## agecat.newC    0.5595      1.787   0.30344    1.0316
## agecat.newD    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
```

```
summary(coxph(formula = Surv(time, censor) ~ agecat+drug, method = "efron"))
print("---")
summary(coxph(formula = Surv(time, censor) ~ agecat.new+drug, method = "efron"))
detach(hmohiv)
```

```
## Call:
## coxph(formula = Surv(time, censor) ~ agecat + drug, method = "efron")
##
##   n= 100, number of events= 80
##
##           coef exp(coef) se(coef)     z Pr(>|z|)
## agecatB 1.2569    3.5146   0.4596 2.735  0.00624 **
## agecatC 1.3194    3.7411   0.4671 2.825  0.00473 **
## agecatD 2.0152    7.5019   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
## agecatB     3.515     0.2845     1.428     8.651
## agecatC     3.741     0.2673     1.498     9.345
## agecatD     7.502     0.1333     2.926    19.234
## drug        2.442     0.4096     1.487     4.009
##
## 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
##
## [1] "---"
## Call:
## coxph(formula = Surv(time, censor) ~ agecat.new + drug, method = "efron")
##
##   n= 100, number of events= 80
##
##                coef exp(coef) se(coef)      z Pr(>|z|)
## agecat.newB -0.7582    0.4685   0.2899 -2.615  0.00891 **
## agecat.newC -0.6958    0.4987   0.3151 -2.208  0.02722 *
## agecat.newD -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
## agecat.newB    0.4685     2.1345   0.26541    0.8269
## agecat.newC    0.4987     2.0053   0.26892    0.9247
## agecat.newD    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
```

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?

Answer: It does not seem to help with respect to the concordance index or the p-values, as they are identical no matter which coding you use. The hazard rations are different, as well as their confidence intervals. This is due to the ratios changing based on the reference group.

Question 4

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

```
## 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
```

```
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
```

```r
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
```

```r
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
```

```r
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
```

Thats a whole lot of output. Summary:

age concordance= 0.532 treat Concordance= 0.543 ndrugtx Concordance= 0.549 hercoc Concordance = 0.523 * different than book all Concordance= 0.585

```
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
```

We get a concordance of 0.587 with only the age, treat and ndrugtx variables.