

Week 7 Assignment

Ben Arancibia

July 19, 2015

Blayney Questions 1-4

- 1) 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.

```
library(survival)
```

```
## Warning: package 'survival' was built under R version 3.1.3
```

```
data(ovarian)
```

```
knitr::kable(ovarian)
```

futime	fustat	age	resid.ds	rx	ecog.ps
59	1	72.3315	2	1	1
115	1	74.4932	2	1	1
156	1	66.4658	2	1	2
421	0	53.3644	2	2	1
431	1	50.3397	2	1	1
448	0	56.4301	1	1	2
464	1	56.9370	2	2	2
475	1	59.8548	2	2	2
477	0	64.1753	2	1	1
563	1	55.1781	1	2	2
638	1	56.7562	1	1	2
744	0	50.1096	1	2	1
769	0	59.6301	2	2	2
770	0	57.0521	2	2	1
803	0	39.2712	1	1	1
855	0	43.1233	1	1	2
1040	0	38.8932	2	1	2
1106	0	44.6000	1	1	1
1129	0	53.9068	1	2	1

futime	fustat	age	resid.ds	rx	ecog.ps
1206	0	44.2055	2	2	1
1227	0	59.5890	1	2	2
268	1	74.5041	2	1	2
329	1	43.1370	2	1	1
353	1	63.2192	1	2	2
365	1	64.4247	2	2	1
377	0	58.3096	1	2	1

Data loaded and now create the plots.

```
my.KMest4 <- survfit(Surv(futime, fustat) ~ rx, data = ovarian)

plot(my.KMest4, main = "OVARIAN CANCER - OVERALL SURVIVAL", col.main = "black",
     xlab = "Time (Days)", ylab = "Overall Survival Proportion", col.lab = "blue",
     cex.lab = 0.9, col = c("red", "blue"), mark.time = TRUE, lty = 2:3)

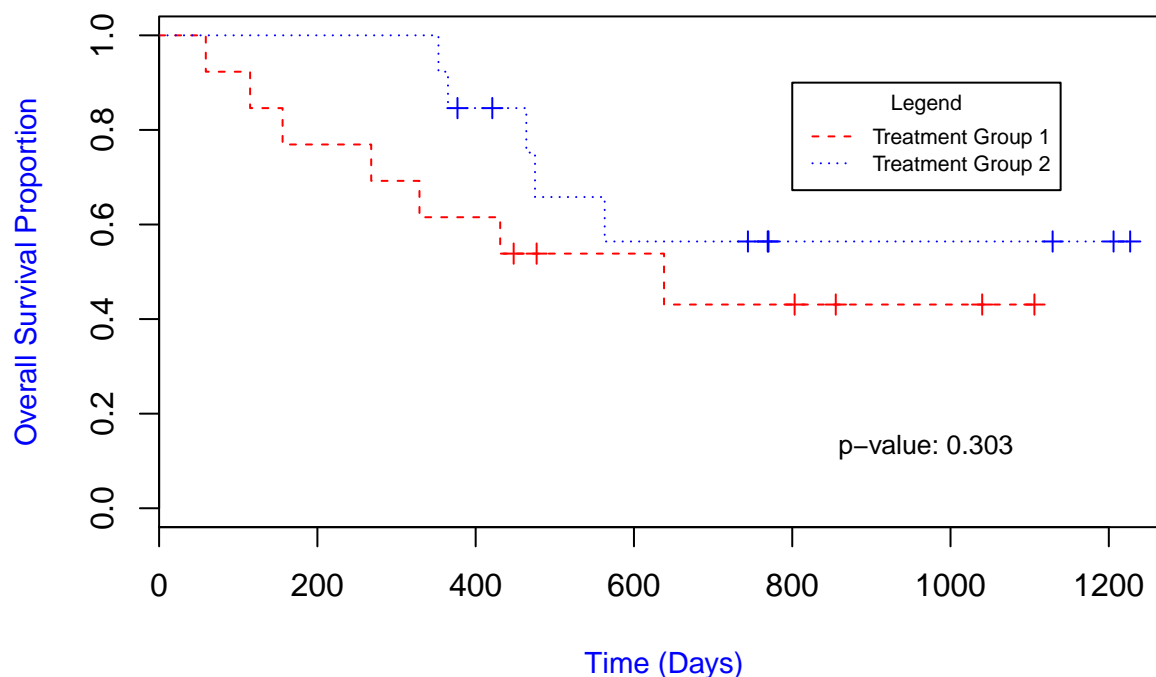
legend(800, 0.9, title = "Legend", c("Treatment Group 1", "Treatment Group 2"),
     lty = 2:3, col = c("red", "blue"), cex = 0.7)

survdifff(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
```

```
legend(800, 0.2, c("p-value: 0.303"), cex = 0.8, box.col = "white")
```

OVARIAN CANCER – OVERALL SURVIVAL



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

```
library(KMsurv)
data(alloauto)
knitr::kable(head(alloauto))
```

time	type	delta
0.030	1	1
0.493	1	1
0.855	1	1
1.184	1	1
1.283	1	1
1.480	1	1

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.

```
my.KMest4 <- survfit(Surv(time, delta) ~ type, data = alloauto, conf.type = "none")
plot(my.KMest4, main = "Leukemia - Free Survival", col.main = "black", xlab = "Time (Months)",
```

```

ylab = "Relapse-Free Survival Proportion", col.lab = "blue", cex.lab = 0.9,
col = c("red", "blue"), mark.time = TRUE, lty = 2:3)

survdif(Surv(time, delta) ~ type, data = alloauto)

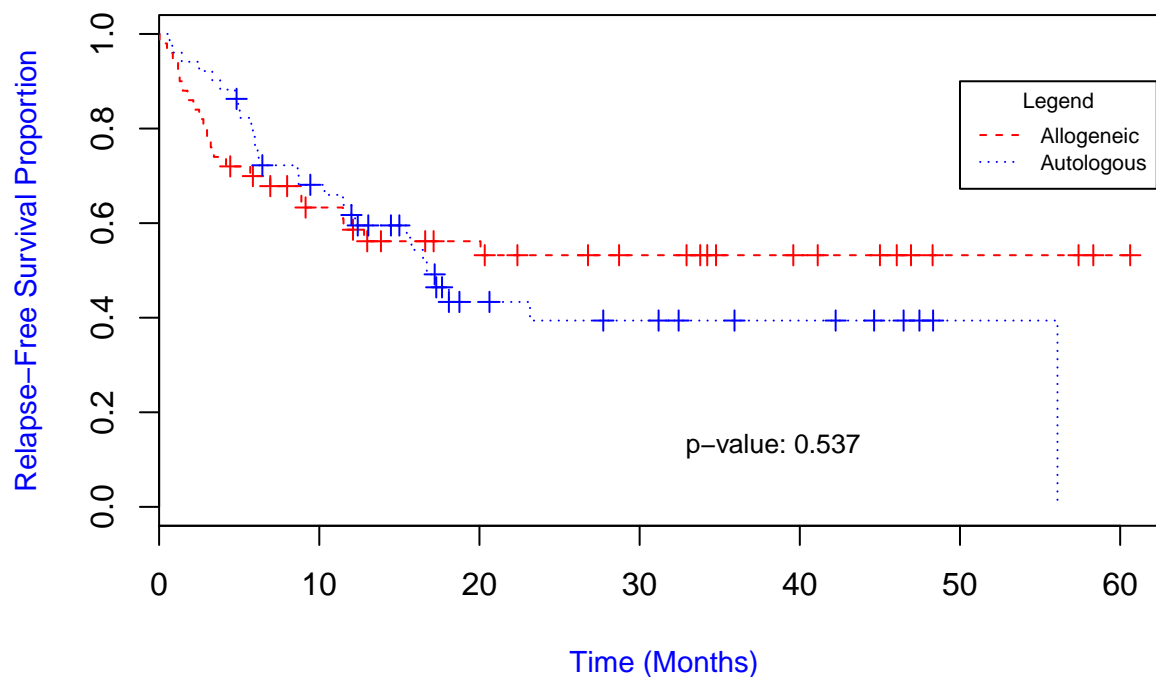
## Call:
## survdiff(formula = Surv(time, delta) ~ type, data = alloauto)
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## type=1  50      22      24.2      0.195      0.382
## type=2  51      28      25.8      0.182      0.382
##
## Chisq= 0.4  on 1 degrees of freedom, p= 0.537

legend(50, 0.9, title = "Legend", c("Allogeneic", "Autologous"), lty = 2:3,
      col = c("red", "blue"), cex = 0.7)

legend(30, 0.2, c("p-value: 0.537"), cex = 0.8, box.col = "white")

```

Leukemia – Free Survival



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

```

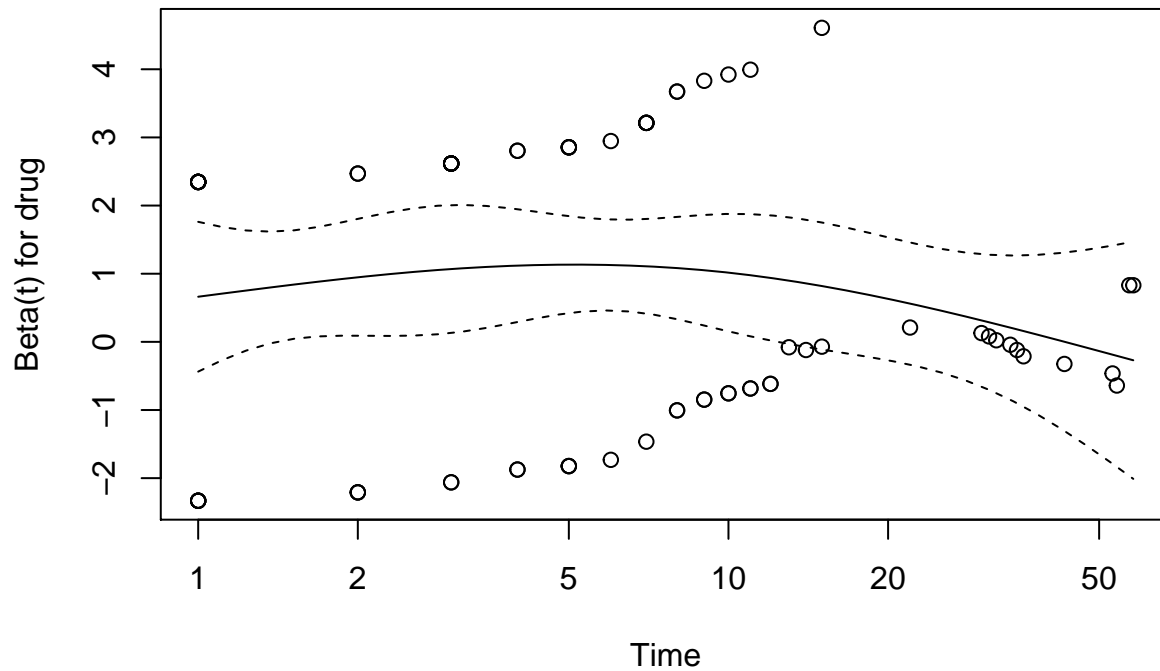
hmohiv <- read.table("http://www.ats.ucla.edu/stat/R/examples/asa/hmohiv.csv",
  sep = ",", header = TRUE)

hmohiv$time <- as.numeric(as.Date(hmohiv[, 7], "%m/%d/%Y") - as.Date(hmohiv[,
  6], "%m/%d/%Y"))
hmohiv$time <- round(hmohiv$time/30.5)

```

```
s_obj <- Surv(hmohiv$time, hmohiv$censor)
# WORK THROUGH EXAMPLE 2.1-2.

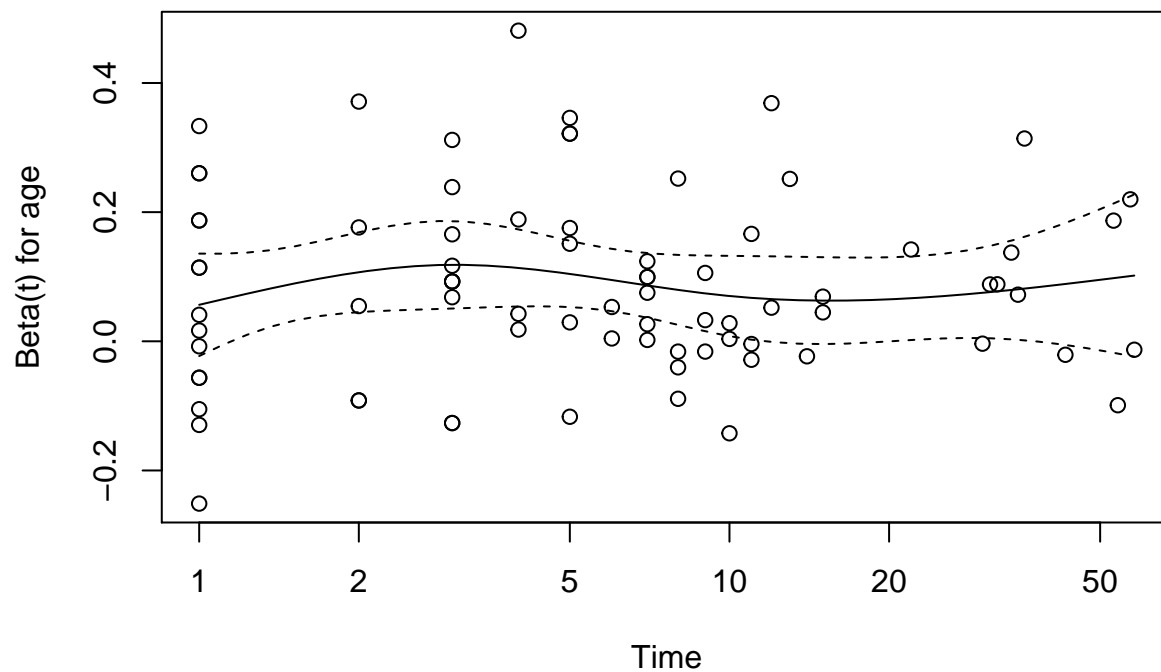
drug.coxph <- coxph(Surv(time, censor) ~ drug, method = "efron", data = hmohiv)
drug_ph <- cox.zph(drug.coxph, transform = "log")
plot(drug_ph[1, ])
```



```
drug_ph
```

```
##          rho chisq      p
## drug -0.0746 0.417 0.518
```

```
age.coxph <- coxph(Surv(time, censor) ~ age, method = "efron", data = hmohiv)
age_ph <- cox.zph(age.coxph, transform = "log")
plot(age_ph[1, ])
```



```
age_ph
```

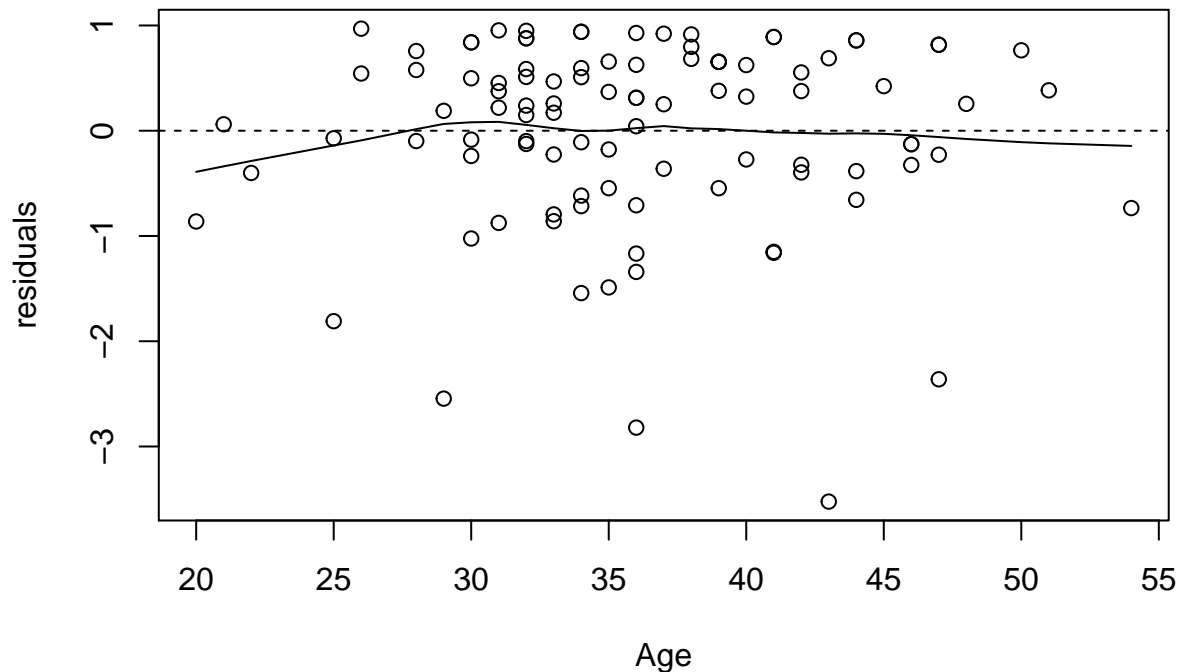
```
##      rho  chisq    p
## age 0.0164 0.0179 0.894
```

```
res <- residuals(age.coxph, type = "martingale")
X <- as.matrix(hmohiv[, "age"]) # matrix of covariates
plot(X[, 1], res, xlab = c("Age")[1], ylab = "residuals")
abline(h = 0, lty = 2) + lines(lowess(X[, 1], res, iter = 0))
```

```
## numeric(0)
```

```
library(car)
```

```
## Warning: package 'car' was built under R version 3.1.3
```



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

```
## Call:
## coxph(formula = Surv(hmohiv$time, hmohiv$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
```

```
agecat <- recode(hmohiv$age, "20:29='D'; 30:34='B'; 35:39='C';40:54='A'", as.factor = T)
```

```
age <- coxph(formula = Surv(hmohiv$time, hmohiv$censor) ~ agecat, method = "efron")
summary(age)
```

```
## Call:
## coxph(formula = Surv(hmohiv$time, hmohiv$censor) ~ agecat, method = "efron")
##
##    n= 100, number of events= 80
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## agecatB -0.7114    0.4909   0.2870 -2.479   0.0132 *
## agecatC -0.5807    0.5595   0.3122 -1.860   0.0629 .
## agecatD -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
## agecatB    0.4909      2.037   0.27971   0.8617
## agecatC    0.5595      1.787   0.30344   1.0316
## agecatD    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
```

```
multi <- coxph(formula = Surv(hmohiv$time, hmohiv$censor) ~ agecat + hmohiv$drug,
  method = "efron")
summary(multi)
```

```
## Call:
## coxph(formula = Surv(hmohiv$time, hmohiv$censor) ~ agecat + hmohiv$drug,
##    method = "efron")
##
##    n= 100, number of events= 80
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## agecatB   -0.7582    0.4685   0.2899 -2.615   0.00891 **
## agecatC   -0.6958    0.4987   0.3151 -2.208   0.02722 *
## agecatD   -2.0152    0.1333   0.4804 -4.195 2.73e-05 ***
## hmohiv$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    0.4685      2.1345   0.26541   0.8269
## agecatC    0.4987      2.0053   0.26892   0.9247
## agecatD    0.1333      7.5019   0.05199   0.3418
## hmohiv$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?

It appears that the pvalues in the univariate and multivariate increase and the larger reference group for age harms. The hazard ratios and confidence intervals really decrease. Concordance values seem to decrease.

4) For this example, we'll look at data involving drug treatment programs.

First, read in the table:

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

```
## The following object is masked _by_ .GlobalEnv:
##
## age
```

```
knitr::kable(head(uis))
```

id	age	becktota	hercoc	ivhx	ndrugtx	race	treat	site	los	time	ensor
1	39	9.00	4	3	1	0	1	0	123	188	1
2	33	34.00	4	2	8	0	1	0	25	26	1
3	33	10.00	2	3	3	0	1	0	7	207	1
4	32	20.00	4	3	1	0	0	0	66	144	1
5	24	5.00	2	1	5	1	1	0	173	551	0
6	30	32.55	3	3	1	0	1	0	16	32	1

For simplicity, we'll only work with the variables mentioned above:

```
uis_small <- uis[, c(1, 2, 4, 6, 8, 9, 11, 12)]
```

For further simplicity, we'll remove the patients with missing values:

```
tiny_uis <- uis_small[apply(uis_small, 1, function(x) !any(is.na(x))), ]
```

We'll assume that the two different sites are different centres, so we'll want to stratify by this variable. You can do this via:

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

Now, repeat the univariate analyses for the variables: treat, age, ndruxt and hercoc. Make a note of the concordance index for each. Create a multivariate model using all four variables. What is the concordance index?

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

```
## Call:
## coxph(formula = Surv(time, censor) ~ treat, data = tiny_uis,
##       method = "efron")
##
##      n= 593, number of events= 481
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## treat -0.23023    0.79435  0.09143 -2.518   0.0118 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## treat    0.7943      1.259    0.664    0.9502
##
## Concordance= 0.537 (se = 0.012 )
## Rsquare= 0.011 (max possible= 1 )
## Likelihood ratio test= 6.34 on 1 df,  p=0.01177
## Wald test            = 6.34 on 1 df,  p=0.0118
## Score (logrank) test = 6.37 on 1 df,  p=0.01161
```

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

```
## Call:
## coxph(formula = Surv(time, censor) ~ age, data = tiny_uis, method = "efron")
##
##      n= 593, number of events= 481
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## age -0.014273    0.985829  0.007343 -1.944   0.0519 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## age    0.9858      1.014    0.9717      1
##
## Concordance= 0.526 (se = 0.014 )
## Rsquare= 0.006 (max possible= 1 )
## Likelihood ratio test= 3.83 on 1 df,  p=0.05039
## Wald test            = 3.78 on 1 df,  p=0.05192
## Score (logrank) test = 3.78 on 1 df,  p=0.05185
```

```
ndrugtx <- coxph(Surv(time, censor) ~ ndrugtx, method = "efron", data = tiny_uis)
summary(ndrugtx)
```

```
## Call:
## coxph(formula = Surv(time, censor) ~ ndrugtx, data = tiny_uis,
##       method = "efron")
##
## n= 593, number of events= 481
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## ndrugtx 0.030712  1.031188 0.007543 4.071 4.67e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## ndrugtx      1.031      0.9698      1.016      1.047
##
## Concordance= 0.545 (se = 0.014 )
## Rsquare= 0.024 (max possible= 1 )
## Likelihood ratio test= 14.33 on 1 df,  p=0.0001533
## Wald test               = 16.58 on 1 df,  p=4.675e-05
## Score (logrank) test = 16.71 on 1 df,  p=4.358e-05
```

```
hercoc <- coxph(Surv(time, censor) ~ hercoc, method = "efron", data = tiny_uis)
summary(hercoc)
```

```
## Call:
## coxph(formula = Surv(time, censor) ~ hercoc, data = tiny_uis,
##       method = "efron")
##
## n= 593, number of events= 481
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## hercoc -0.07720   0.92570 0.04164 -1.854  0.0637 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## hercoc      0.9257      1.08      0.8532      1.004
##
## Concordance= 0.525 (se = 0.014 )
## Rsquare= 0.006 (max possible= 1 )
## Likelihood ratio test= 3.4 on 1 df,  p=0.06523
## Wald test              = 3.44 on 1 df,  p=0.06372
## Score (logrank) test = 3.44 on 1 df,  p=0.06351
```

Concordances

treat = .537

age = 0.526

ndrugtx = .545

hercoc = .525

```
# multivariant
```

```
fit_four_cox_model <- coxph(Surv(time, censor) ~ hercoc + treat + age + ndrugtx,  
  method = "efron", data = tiny_uis)
```

```
summary(fit_four_cox_model)
```

```
## Call:  
## coxph(formula = Surv(time, censor) ~ hercoc + treat + age + ndrugtx,  
##       data = tiny_uis, method = "efron")  
##  
##    n= 593, number of events= 481  
##  
##              coef exp(coef)  se(coef)      z Pr(>|z|)  
## hercoc -0.063348  0.938617  0.043388 -1.460  0.14428  
## treat  -0.232669  0.792416  0.091776 -2.535  0.01124 *  
## age    -0.024750  0.975554  0.007688 -3.219  0.00129 **  
## ndrugtx 0.035494  1.036131  0.007827  4.535 5.76e-06 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
##              exp(coef) exp(-coef) lower .95 upper .95  
## hercoc      0.9386      1.0654      0.8621      1.0219  
## treat       0.7924      1.2620      0.6620      0.9486  
## age         0.9756      1.0251      0.9610      0.9904  
## ndrugtx     1.0361      0.9651      1.0204      1.0521  
##  
## Concordance= 0.575 (se = 0.014 )  
## Rsquare= 0.053 (max possible= 1 )  
## Likelihood ratio test= 32.37 on 4 df,  p=1.611e-06  
## Wald test              = 34.9 on 4 df,  p=4.878e-07  
## Score (logrank) test = 35 on 4 df,  p=4.645e-07
```

multi concordance = 0.575

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

```
library(MASS)
```

```
## Warning: package 'MASS' was built under R version 3.1.3
```

```
stepAIC(fit_four_cox_model)
```

```
## Start:  AIC=5521.79  
## Surv(time, censor) ~ hercoc + treat + age + ndrugtx  
##  
##              Df      AIC  
## <none>          5521.8  
## - hercoc       1 5521.9  
## - treat        1 5526.2  
## - age          1 5530.3  
## - ndrugtx      1 5537.3
```

```
## Call:
## coxph(formula = Surv(time, censor) ~ hercoc + treat + age + ndrugh,
##       data = tiny_uis, method = "efron")
##
##
##              coef exp(coef) se(coef)      z      p
## hercoc  -0.06335   0.93862  0.04339 -1.46  0.1443
## treat   -0.23267   0.79242  0.09178 -2.54  0.0112
## age     -0.02475   0.97555  0.00769 -3.22  0.0013
## ndrugh  0.03549   1.03613  0.00783  4.54 5.8e-06
##
## Likelihood ratio test=32.4 on 4 df, p=1.61e-06
## n= 593, number of events= 481
```

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

Variables retained are: Ndrugh, Age, Treat

```
three <- coxph(Surv(time, censor) ~ +treat + age + ndrugh, method = "efron",
              data = tiny_uis)

summary(three)
```

```
## Call:
## coxph(formula = Surv(time, censor) ~ +treat + age + ndrugh,
##       data = tiny_uis, method = "efron")
##
##      n= 593, number of events= 481
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## treat   -0.243286  0.784047  0.091483 -2.659  0.00783 **
## age     -0.022886  0.977374  0.007580 -3.019  0.00253 **
## ndrugh  0.037147  1.037845  0.007668  4.844 1.27e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## treat      0.7840      1.2754    0.6553    0.938
## age       0.9774      1.0232    0.9630    0.992
## ndrugh     1.0378      0.9635    1.0224    1.054
##
## Concordance= 0.577 (se = 0.014 )
## Rsquare= 0.05 (max possible= 1 )
## Likelihood ratio test= 30.25 on 3 df, p=1.224e-06
## Wald test = 33.13 on 3 df, p=3.023e-07
## Score (logrank) test = 33.25 on 3 df, p=2.849e-07
```

Ndrugh, Age, Treat: Concordance= 0.577