# 432 Homework 6 Answer Sketch

*432 Staff*

*Due 2018-04-22 at 2 pm. Version: 2019-03-18*

```
library(skimr); library(rms); library(ggrepel)
library(readxl); library(survival); library(OIsurv)
library(survminer); library(broom); library(tidyverse)
```

```
remission <- read.csv("remission.csv")%>% tbl_df
umaru <- read.csv("umaru.csv") %>% tbl_df
```

## 1 Question 1 (15 points)

Create a visualization (using R) of real data on a subject that is meaningful to you, and share it (the visualization (and the code you used to build it) with us. The visualization should be of a professional quality, include proper labels and a title, as well as a caption of no more than 50 words that highlights the key result. If you decide to create a new visualization based on a revision of someone else's work, you must share with us that original work, as well.

We will grade Question 1 strictly based on the quality of the visualization, its title and caption, in terms of being attractive, well-labeled and useful for representing the data, and how well it adheres to general principles for good visualizations we've seen in 431 and 432.

## 2 Question 2 (20 points)

Write an essay (between 150 and 300 words) describing the background, creation and meaning of the visualization you created in Question 1, providing us with the context we need to understand why this is a meaningful, and perhaps important visualization. In your short description, address each of the following issues.

- How does this visualization help its audience understand the world better?
- Why is this particular visualization effective, and what are the design features it uses that we can learn from to help us make more effective visualizations?
- How is this visualization coded in R? What tools did you use, and why did you select them?

**This is an essay question. We don't write answer sketches for essays.**

### 2.1 Setup for Questions 3-5

The `umaru.csv` data file contains information for 575 subjects selected from the UMARU IMPACT study collaborative project done by the University of Massachusetts AIDS Research Unit over 5 years (1989-1994). Various versions of this data set are frequently used in survival analysis texts. I've tweaked your data set enough that you'll see some different results. The study included two concurrent randomized trials of residential treatment for drug abuse. The key question is to compare treatment programs of different planned durations in terms of their ability to reduce drug abuse and prevent high-risk HIV behavior. Here's a codebook:

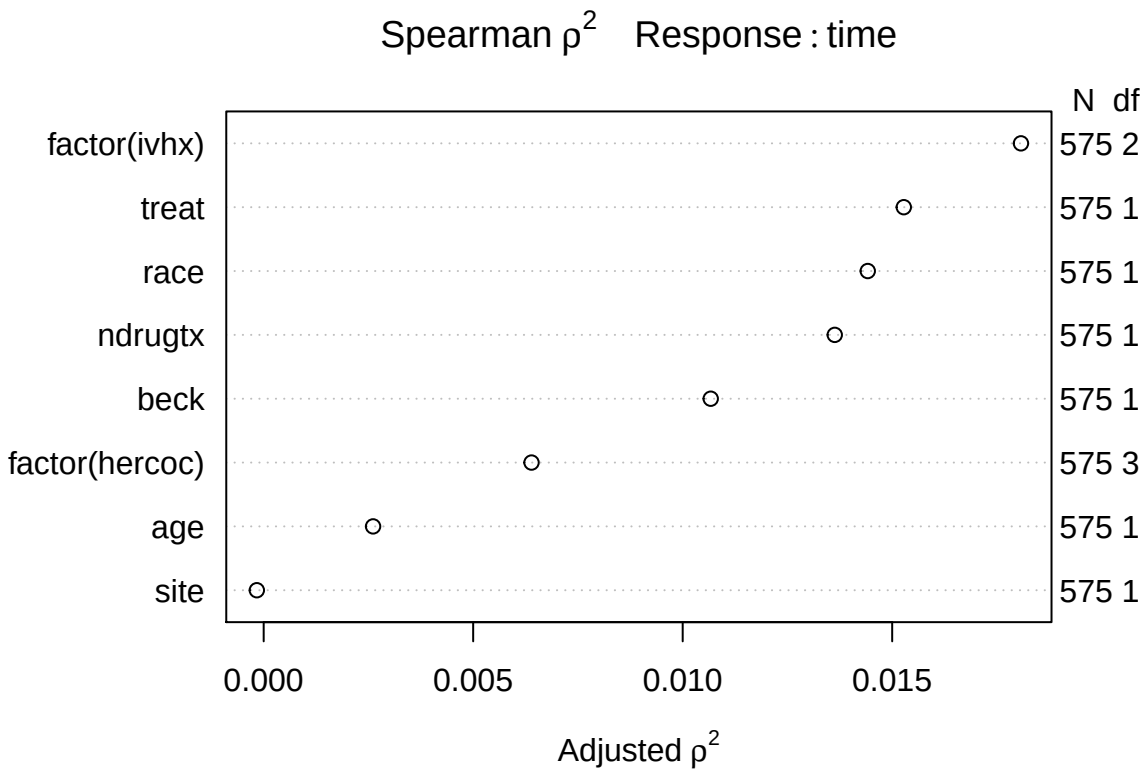
| Variable | Description |
|---|---|

| Variable | Description |
|---|---|
| subject | Subject ID #, ranging from 1001 - 1575 |
| age | age at enrollment, in years |
| beck | Beck Depression Score at admission |
| hercoc | heroin or cocaine use during the 3 months prior to admission (1 = Heroin & Cocaine, 2 = Heroin only, 3 = Cocaine only, 4 = Neither Heroin nor Cocaine) |
| ivhx | IV drug use history at admission (1 = never, 2 = previous but not recent, 3 = recent) |
| ndrugtx | # of prior drug treatments |
| race | subject's race (0 = white, 1 = other) |
| treat | treatment randomization assignment (Long, or Short) |
| site | treatment site (A or B) |
| lot | Length of Treatment (Exit Date - Admission Date), in days |
| time | Time to Return to Drug Use (measured from Admission Date), in days |
| censor | Returned to Drug Use indicator (1 = returned to drug use, 0 = otherwise) |

# 3 Question 3 (15 points)

Build a Cox model, using `treat` as a predictor, and spending degrees of freedom in any way you like with the rest of the available predictors (i.e. *everything but* `subject`, `lot`, `time` and `censor`) in the data set, so long as you do not exceed a total of 12 degrees of freedom, predicting the time to return to drug use. You'll probably want to use a Spearman rho-squared plot to make your selection, in which case you should stick with the model you develop using that tool, regardless of its eventual performance. Specify your model carefully, and interpret the hazard ratio for `treat` implied by your new model.

```
plot(spearman2(time ~ treat + age + beck + factor(hercoc) +
       factor(ivhx) + ndrugtx + race + site, data=umaru))
```




The key variables look like:

- `ivhx`, which is a three-level categorical variable, which I'll treat as nominal here (but you could certainly have chosen to treat it as ordinal, too)
- `treat`, of course, which is binary
- `race`, which is binary here
- `ndrugtx`, which is continuous
- `beck`, which is also continuous
- and maybe the `hercoc` indicator, which is a four-level categorical variable

If we include those six variables, that chews up 6 degrees of freedom, so in terms of adding non-linearity, we have about 5-6 available degrees of freedom to use. I chose to look at the interaction of ivhx and treat, and of ivhx and race, and then add a restricted cubic spline with 3 knots for ndrugtx, to spend my remaining degrees of freedom. You probably chose some other strategy, which is fine, so long as you stayed to the limit of 12 degrees of freedom spent.

## 3.1 Using `cph` to fit Model A

```
umaru$ivhx <- factor(umaru$ivhx)
umaru$hercoc <- factor(umaru$hercoc)
d <- datadist(umaru)
options(datadist="d")

mod_A <- cph(Surv(time, censor==1) ~ ivhx + treat + race +
               ivhx*treat + rcs(ndrugtx, 3) + beck + hercoc,
             data=umaru, x=TRUE, y=TRUE, surv=TRUE)
```

```
mod_A
```

```
Cox Proportional Hazards Model

 cph(formula = Surv(time, censor == 1) ~ ivhx + treat + race +
     ivhx * treat + rcs(ndrugtx, 3) + beck + hercoc, data = umaru,
     x = TRUE, y = TRUE, surv = TRUE)

                    Model Tests       Discrimination
                                         Indexes
 Obs       575    LR chi2     35.75    R2       0.060
 Events    464    d.f.           12    Dxy      0.172
 Center 0.4742    Pr(> chi2) 0.0004    g        0.324
                  Score chi2  37.36    gr       1.382
                  Pr(> chi2) 0.0002

                      Coef    S.E.   Wald Z Pr(>|Z|)
 ivhx=2                0.1779 0.1866  0.95  0.3405
 ivhx=3                0.1647 0.1776  0.93  0.3537
 treat=Short           0.2604 0.1562  1.67  0.0955
 race                 -0.2165 0.1139 -1.90  0.0573
 ndrugtx               0.0403 0.0402  1.00  0.3161
 ndrugtx'             -0.0211 0.0522 -0.40  0.6863
 beck                  0.0092 0.0049  1.87  0.0612
 hercoc=2              0.0503 0.1503  0.33  0.7378
 hercoc=3             -0.0521 0.1668 -0.31  0.7549
 hercoc=4              0.0329 0.1629  0.20  0.8401
 ivhx=2 * treat=Short -0.2345 0.2674 -0.88  0.3805
 ivhx=3 * treat=Short  0.0245 0.2104  0.12  0.9074
```

```
anova(mod_A)
```

```
               Wald Statistics          Response: Surv(time, censor == 1)

 Factor                                     Chi-Square d.f. P
 ivhx  (Factor+Higher Order Factors)         2.64       4   0.6190
  All Interactions                           1.06       2   0.5882
 treat  (Factor+Higher Order Factors)        6.89       3   0.0754
  All Interactions                           1.06       2   0.5882
 race                                        3.62       1   0.0573
 ndrugtx                                     8.33       2   0.0155
  Nonlinear                                  0.16       1   0.6863
 beck                                        3.51       1   0.0612
 hercoc                                      0.64       3   0.8870
 ivhx * treat  (Factor+Higher Order Factors) 1.06       2   0.5882
 TOTAL NONLINEAR + INTERACTION               1.27       3   0.7352
 TOTAL                                      36.62      12   0.0003
```

```
exp(coef(mod_A)) # to get treat hazard ratio
```

```
            ivhx=2              ivhx=3         treat=Short
         1.1946710           1.1790700           1.2973972
              race             ndrugtx            ndrugtx'
         0.8053370           1.0411148           0.9791435
```

```
              beck             hercoc=2             hercoc=3
         1.0092899            1.0515808            0.9492461
          hercoc=4 ivhx=2 * treat=Short ivhx=3 * treat=Short
         1.0334134            0.7909796            1.0247858
```

```
exp(confint(mod_A)) # 95% CI for hazard ratios
```

```
                         2.5 %   97.5 %
ivhx=2                0.8286908 1.722281
ivhx=3                0.8324332 1.670051
treat=Short           0.9552737 1.762050
race                  0.6442525 1.006698
ndrugtx               0.9622459 1.126448
ndrugtx'              0.8839431 1.084597
beck                  0.9995676 1.019107
hercoc=2              0.7833267 1.411700
hercoc=3              0.6844945 1.316399
hercoc=4              0.7508828 1.422250
ivhx=2 * treat=Short 0.4683569 1.335837
ivhx=3 * treat=Short 0.6784824 1.547846
```

### 3.2 Using coxph to fit Model A

```
mod_A1 <-with(umaru,
              coxph(Surv(time, censor==1) ~
                        ivhx + treat + race +  ivhx*treat + rcs(ndrugtx, 3) + beck + hercoc))

summary(mod_A1)
```

```
Call:
coxph(formula = Surv(time, censor == 1) ~ ivhx + treat + race +
    ivhx * treat + rcs(ndrugtx, 3) + beck + hercoc)

  n= 575, number of events= 464

                             coef exp(coef)  se(coef)      z Pr(>|z|)
ivhx2                    0.177871  1.194671  0.186625  0.953   0.3405
ivhx3                    0.164725  1.179069  0.177620  0.927   0.3537
treatShort               0.260360  1.297397  0.156185  1.667   0.0955 .
race                    -0.216494  0.805337  0.113864 -1.901   0.0573 .
rcs(ndrugtx, 3)ndrugtx   0.040294  1.041117  0.040193  1.003   0.3161
rcs(ndrugtx, 3)ndrugtx' -0.021081  0.979139  0.052188 -0.404   0.6862
beck                     0.009247  1.009290  0.004939  1.872   0.0612 .
hercoc2                  0.050295  1.051581  0.150258  0.335   0.7378
hercoc3                 -0.052087  0.949246  0.166833 -0.312   0.7549
hercoc4                  0.032867  1.033413  0.162948  0.202   0.8402
ivhx2:treatShort        -0.234483  0.790980  0.267373 -0.877   0.3805
ivhx3:treatShort         0.024485  1.024787  0.210402  0.116   0.9074
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

                        exp(coef) exp(-coef) lower .95 upper .95
ivhx2                      1.1947     0.8371    0.8287     1.722
ivhx3                      1.1791     0.8481    0.8324     1.670
```

```
treatShort                 1.2974     0.7708    0.9553     1.762
race                       0.8053     1.2417    0.6443     1.007
rcs(ndrugtx, 3)ndrugtx     1.0411     0.9605    0.9622     1.126
rcs(ndrugtx, 3)ndrugtx'    0.9791     1.0213    0.8839     1.085
beck                       1.0093     0.9908    0.9996     1.019
hercoc2                    1.0516     0.9509    0.7833     1.412
hercoc3                    0.9492     1.0535    0.6845     1.316
hercoc4                    1.0334     0.9677    0.7509     1.422
ivhx2:treatShort           0.7910     1.2643    0.4684     1.336
ivhx3:treatShort           1.0248     0.9758    0.6785     1.548

Concordance= 0.586  (se = 0.014 )
Rsquare= 0.06   (max possible= 1 )
Likelihood ratio test= 35.75  on 12 df,   p=4e-04
Wald test            = 36.62  on 12 df,   p=3e-04
Score (logrank) test = 37.36  on 12 df,   p=2e-04
```

Model A doesn't actually fit very well. Neither the `ivhx` term, nor its interaction with `treat` is of much use, which is a bit of a surprise. The hazard ratio estimate for the "Short" duration `treat`ment is 1.30, but has a 95% CI of (0.96, 1.76) after adjusting for these predictors.

# 4 Question 4 (10 points)

Apply a Cox regression model to predict the time to return to drug use (incorporating censoring appropriately) using the information in `treat`, plus main effects of `age`, `beck`, `site`, `ivhx` and `ndrugtx`. Interpret the meaning of the hazard ratio for `treat`, after adjusting for the other five predictors.

## 4.1 Using `cph` to fit Model B

```
d <- datadist(umaru)
options(datadist="d")
mod_B <- cph(Surv(time, censor==1) ~
               treat + age + beck + site + ivhx + ndrugtx,
             data=umaru, x=TRUE, y=TRUE, surv=TRUE)
mod_B
```

```
Cox Proportional Hazards Model

 cph(formula = Surv(time, censor == 1) ~ treat + age + beck +
     site + ivhx + ndrugtx, data = umaru, x = TRUE, y = TRUE,
     surv = TRUE)

                      Model Tests       Discrimination
                                           Indexes
 Obs        575    LR chi2     43.24    R2       0.072
 Events     464    d.f.            7    Dxy      0.183
 Center -0.3972    Pr(> chi2) 0.0000    g        0.354
                   Score chi2  44.77    gr       1.425
                   Pr(> chi2) 0.0000

              Coef    S.E.   Wald Z Pr(>|Z|)
 treat=Short  0.2385 0.0936  2.55  0.0109
```

```
 age         -0.0297 0.0082 -3.64  0.0003
 beck         0.0072 0.0050  1.46  0.1453
 site=B      -0.0595 0.1061 -0.56  0.5746
 ivhx=2       0.2215 0.1365  1.62  0.1048
 ivhx=3       0.3852 0.1166  3.30  0.0010
 ndrugtx      0.0290 0.0083  3.51  0.0005
```

```
exp(coef(mod_B)) # to get treat hazard ratio
```

```
treat=Short         age        beck      site=B      ivhx=2      ivhx=3
  1.2694070   0.9707600   1.0072695   0.9421917   1.2479469   1.4699775
    ndrugtx
  1.0294010
```

```
exp(confint(mod_B)) # 95% CI for hazard ratios
```

```
                2.5 %    97.5 %
treat=Short 1.0565576 1.5251361
age         0.9553541 0.9864144
beck        0.9974982 1.0171365
site=B      0.7653109 1.1599537
ivhx=2      0.9549224 1.6308881
ivhx=3      1.1696938 1.8473500
ndrugtx     1.0128655 1.0462064
```

## 4.2 Using coxph to fit Model B

```
mod_B1 <-with(umaru, coxph(Surv(time, censor==1) ~
                treat  + age + beck + site + ivhx +ndrugtx))

summary(mod_B1)
```

```
Call:
coxph(formula = Surv(time, censor == 1) ~ treat + age + beck +
    site + ivhx + ndrugtx)

  n= 575, number of events= 464

                coef exp(coef)  se(coef)      z Pr(>|z|)
treatShort  0.238550  1.269407  0.093641  2.547 0.010850 *
age        -0.029676  0.970760  0.008162 -3.636 0.000277 ***
beck        0.007243  1.007269  0.004974  1.456 0.145303
siteB      -0.059547  0.942191  0.106087 -0.561 0.574592
ivhx2       0.221501  1.247948  0.136546  1.622 0.104767
ivhx3       0.385248  1.469979  0.116586  3.304 0.000952 ***
ndrugtx     0.028976  1.029400  0.008262  3.507 0.000453 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

           exp(coef) exp(-coef) lower .95 upper .95
treatShort    1.2694     0.7878    1.0566    1.5251
age           0.9708     1.0301    0.9554    0.9864
beck          1.0073     0.9928    0.9975    1.0171
```

```
siteB         0.9422     1.0614    0.7653    1.1600
ivhx2         1.2479     0.8013    0.9549    1.6309
ivhx3         1.4700     0.6803    1.1697    1.8474
ndrugtx       1.0294     0.9714    1.0129    1.0462

Concordance= 0.592  (se = 0.014 )
Rsquare= 0.072   (max possible= 1 )
Likelihood ratio test= 43.24  on 7 df,   p=3e-07
Wald test            = 44.38  on 7 df,   p=2e-07
Score (logrank) test = 44.77  on 7 df,   p=2e-07
```

The hazard ratio estimate for `treat=Short`, after adjusting for the other five predictors, is 1.27, and we have a 95% confidence interval ranging from (1.06, 1.52). Since this interval completely exceeds 1, the hazard function is statistically significantly larger (i.e. worse, at the 5% significance level) for those randomized to the Short treatment than for those randomized to the Long duration treatment.

# 5 Question 5 (20 points)

Compare the two models you have fit in Questions 3 and 4, specifying which one you prefer and why. Be sure to include both a comparison of the quality of fit from each model (be sure to at least two ways to assess that quality of fit), and an assessment of adherence to the assumptions of a proportional hazards model for your final selection. Validate the summary statistics describing your chosen model, and explain what those results mean, too.

## 5.1 Checking the Proportional Hazards Assumption

```
(modA_1ph <-cox.zph(mod_A1, transform="km", global=TRUE))
```

```
                           rho  chisq      p
ivhx2                  -0.0226  0.234 0.6285
ivhx3                  -0.0209  0.197 0.6572
treatShort             -0.0923  3.941 0.0471
race                    0.0529  1.297 0.2547
rcs(ndrugtx, 3)ndrugtx  0.0439  1.016 0.3134
rcs(ndrugtx, 3)ndrugtx' -0.0417  0.918 0.3379
beck                   -0.0839  3.032 0.0816
hercoc2                 0.0323  0.493 0.4827
hercoc3                 0.0201  0.191 0.6623
hercoc4                 0.0154  0.112 0.7383
ivhx2:treatShort        0.0508  1.193 0.2748
ivhx3:treatShort        0.0487  1.121 0.2897
GLOBAL                      NA 10.610 0.5626
```

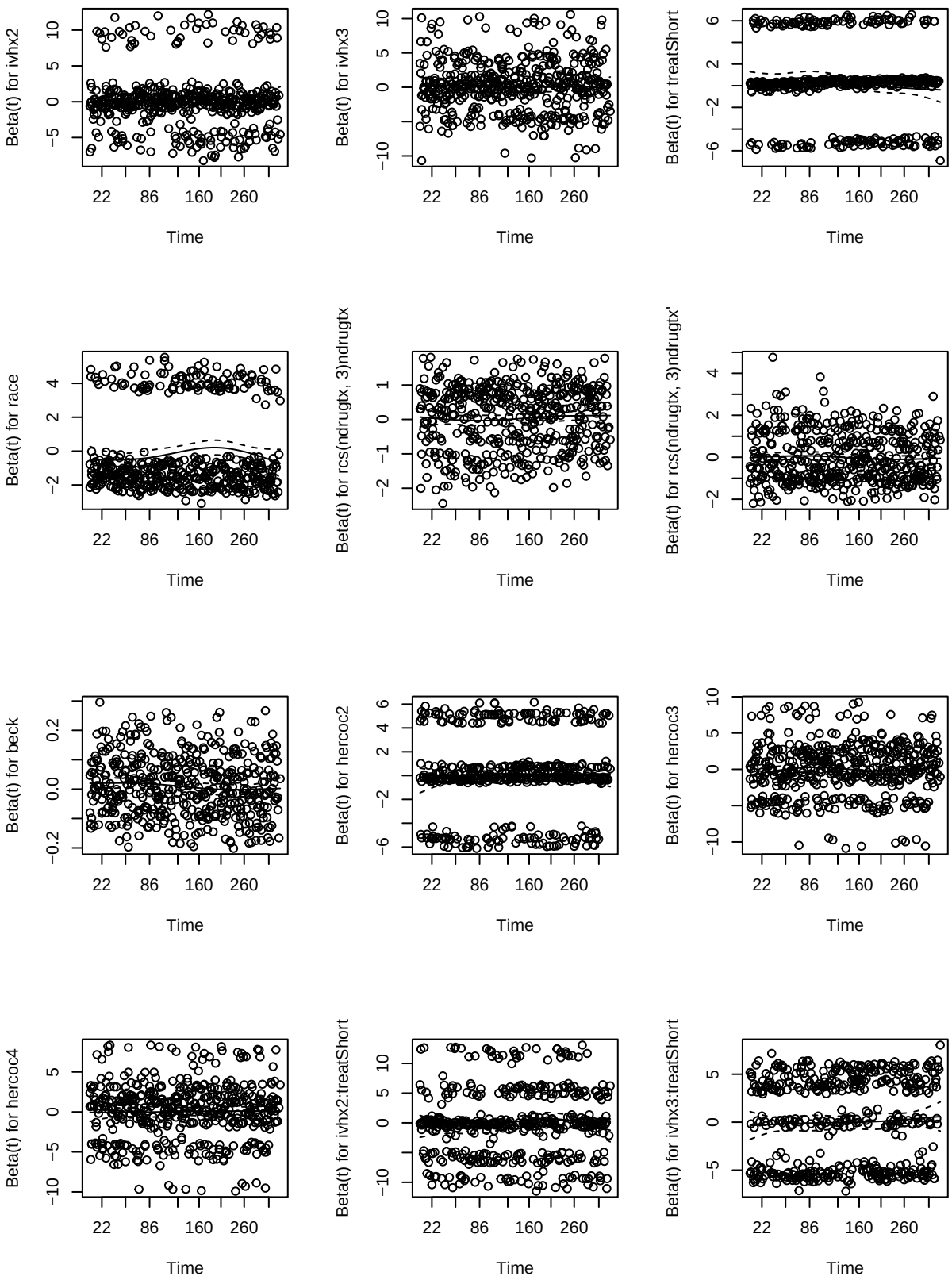
```
(modB_1ph <-cox.zph(mod_B1, transform="km", global=TRUE))
```

```
               rho  chisq      p
treatShort -0.09058 3.8173 0.0507
age         0.02868 0.3604 0.5483
beck       -0.07439 2.4245 0.1194
siteB       0.01873 0.1688 0.6812
ivhx2       0.01764 0.1414 0.7069
ivhx3       0.00507 0.0115 0.9147
```

```
ndrugtx     0.01982 0.1754 0.6754
GLOBAL           NA 7.6137 0.3679
```

#### 5.1.1 Model A from Question 3

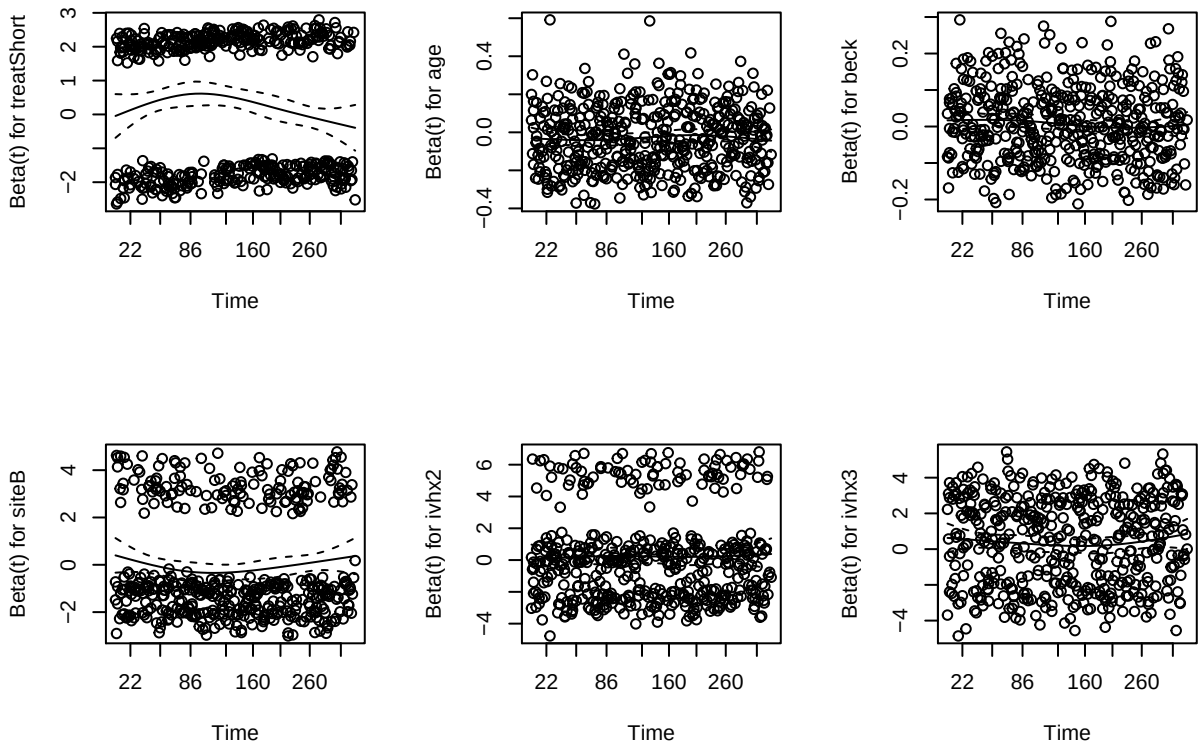
```
par(mfrow=c(2,3))
plot(modA_1ph)
```


Beta(t) for ivhx2
Beta(t) for ivhx3
Beta(t) for treatShort
Beta(t) for race
Beta(t) for rcs(ndrugtx, 3)ndrugtx
Beta(t) for rcs(ndrugtx, 3)ndrugtx'
Beta(t) for beck
Beta(t) for hercoc2
Beta(t) for hercoc3
Beta(t) for hercoc4
Beta(t) for ivhx2:treatShort
Beta(t) for ivhx3:treatShort
Time
22
86
160
260

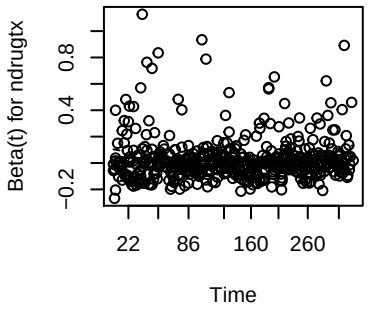
```
par(mfrow=c(1,1))
```

### 5.1.2 Model B from Question 4

```
par(mfrow=c(2,3))
plot(modB_1ph)
```



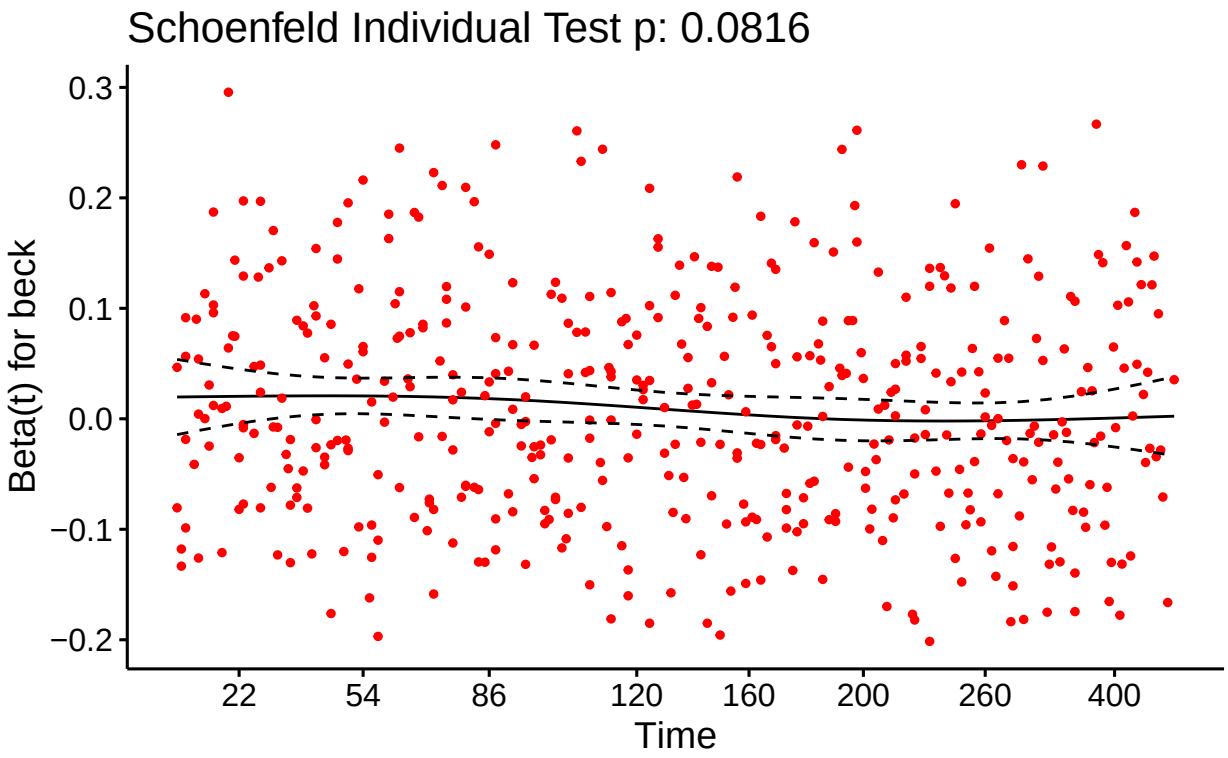

```
par(mfrow=c(1,1))
```




## 5.2 Using the survminer package to check proportional hazards

### 5.2.1 Model A from Question 3

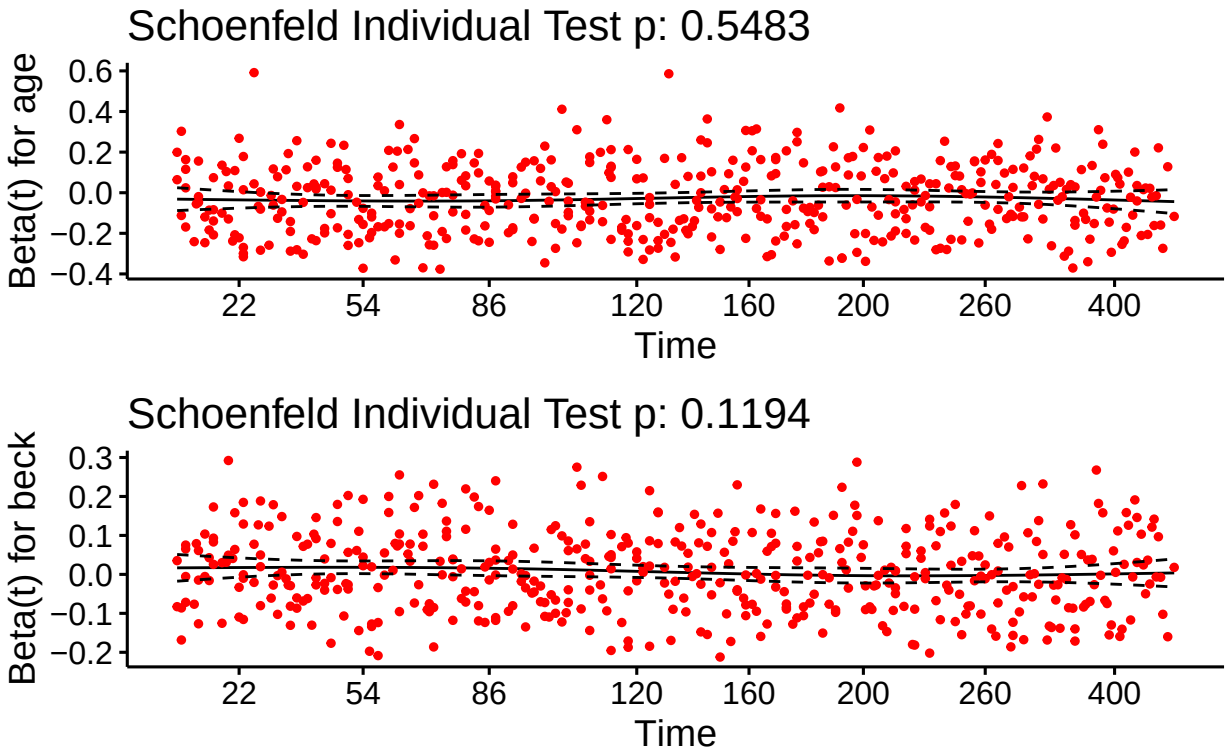
```
#Specifying only continuous variables, this plot will give you the Schoenfeld individual test pvalue ab
ggcoxzph(modA_1ph,  var = c("beck"))
```




#### 5.2.2 Model B from Question 4

```
#Specifying only continuous variables, this plot will give you the Schoenfeld individual test pvalue ab
ggcoxzph(modB_1ph,  var = c("age", "beck"))
```

Global Schoenfeld Test p: 0.3679




## 5.3 Summarizing the Fit

### 5.3.1 Quality of Fit Summaries

```
anova(mod_A)
```

```
                Wald Statistics          Response: Surv(time, censor == 1)

 Factor                                     Chi-Square d.f. P
 ivhx  (Factor+Higher Order Factors)         2.64       4   0.6190
  All Interactions                           1.06       2   0.5882
 treat  (Factor+Higher Order Factors)        6.89       3   0.0754
  All Interactions                           1.06       2   0.5882
 race                                        3.62       1   0.0573
 ndrugtx                                     8.33       2   0.0155
  Nonlinear                                  0.16       1   0.6863
 beck                                        3.51       1   0.0612
 hercoc                                      0.64       3   0.8870
 ivhx * treat  (Factor+Higher Order Factors) 1.06       2   0.5882
 TOTAL NONLINEAR + INTERACTION               1.27       3   0.7352
 TOTAL                                      36.62      12   0.0003
```

```
anova(mod_B)
```

```
                Wald Statistics          Response: Surv(time, censor == 1)
```

```
 Factor     Chi-Square d.f. P
 treat       6.49      1    0.0109
 age        13.22      1    0.0003
 beck        2.12      1    0.1453
 site        0.32      1    0.5746
 ivhx       10.93      2    0.0042
 ndrugtx    12.30      1    0.0005
 TOTAL      44.39      7    <.0001
```

```
AIC(mod_A)
```

```
[1] 5314.551
```

```
AIC(mod_B)
```

```
[1] 5297.064
```

I believe I prefer Model B of these two, based on its superior $R^2$ (although neither is terrific), Somers' d statistic, AIC and BIC. Neither model shows a serious deviation from the proportional hazards assumption, although the $p$ value for the `treat` variable is close to a significant departure from assumptions in Model B.

| Model | $R^2$ | Somers' d | LR test | AIC | BIC | Cox PH global test |
|---|---|---|---|---|---|---|
| A | 0.060 | 0.172 | $p = 0.0004$ | 5315 | 5367 | $p = 0.56$ |
| B | 0.072 | 0.183 | $p < 0.0001$ | 5297 | 5328 | $p = 0.37$ |

On the basis of the ANOVA tables, it looks like a better fit to the data can perhaps be achieved by removing `beck` and `site` from Model B but keeping the other predictors, and perhaps adding back in `race` from Model A, but I didn't see that in the Spearman plot, so I didn't do it.

# 6 Question 6 (20 points)

The `remission.csv` file contains contains initial remission times, in days, for 44 leukemia patients who were randomly allocated to two different treatments, labeled A and B. Some patients were right-censored before their remission times could be fully determined, as indicated by values of `censored` = 1 in the data set. It's worth emphasizing that shorter times to remission indicate better news.

Your task is to plot and compare appropriate estimates of the survival functions for the two treatments, including at least a Kaplan-Meier estimate and a log rank test. Compare median and (restricted) mean survival times appropriately. Write a complete sentence (or several) to accompany each of your estimates and plots. Do not use a regression model.

## 6.1 Compare the Kaplan-Meier estimated survival functions for the two treatments.

```
## establish survival object
## event occurs when censored = 0, but not when censored = 1
remsurv <- with(remission, Surv(time = time, event = 1-censored))

## original data shows third subject is censored
head(remission)
```

```
# A tibble: 6 x 4
  subject treatment  time censored
    <int> <fct>     <int>    <int>
1       1 A            41        0
2       2 A            37        0
3       3 B           217        1
4       4 A           195        0
5       5 B           169        0
6       6 B           103        0
```

```
head(remsurv) ## third patient correctly indicated as censored
```

```
[1]  41   37  217+ 195  169  103
```

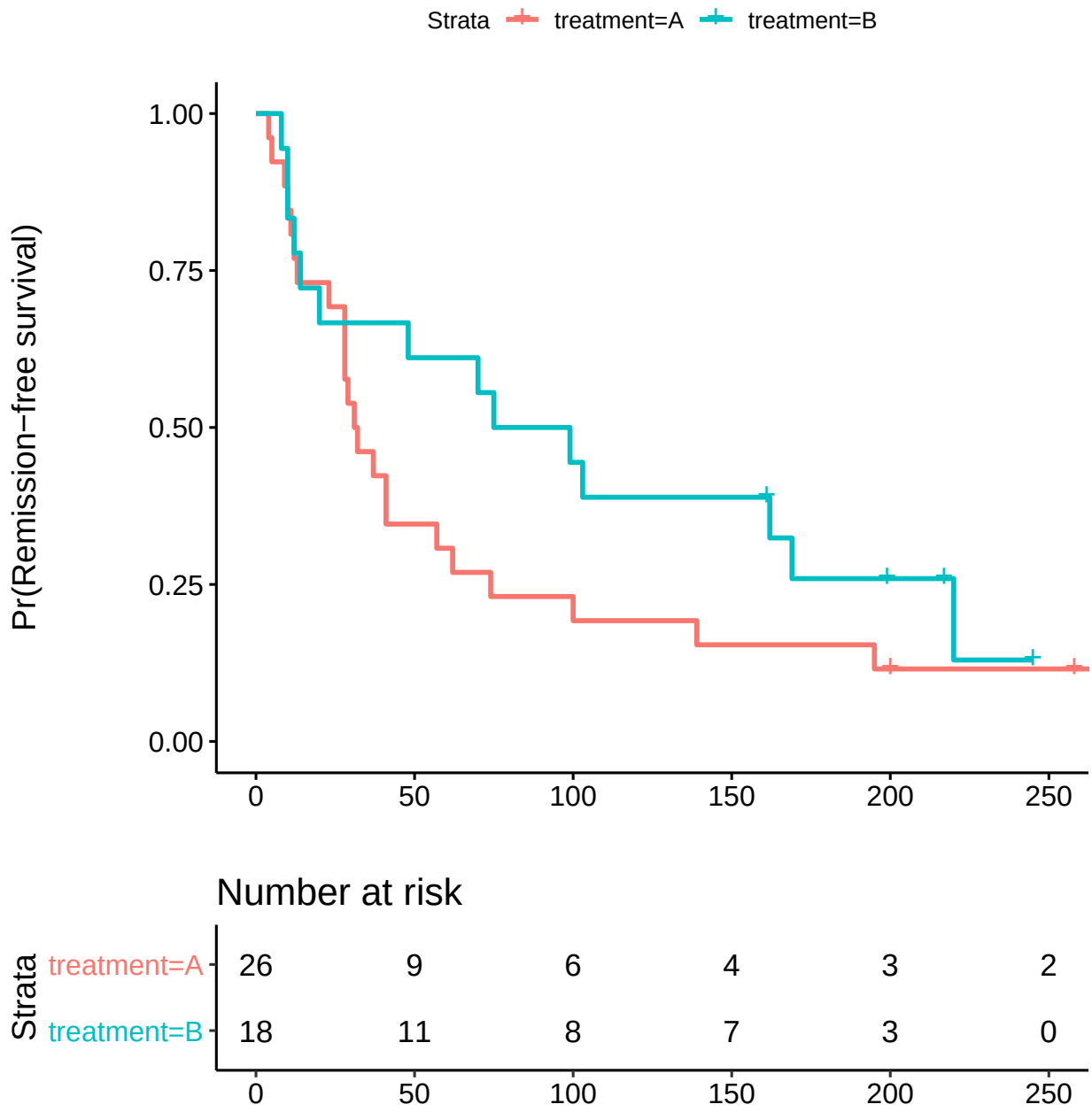
```
remfit <- survfit(remsurv ~ remission$treatment)
print(remfit, print.rmean=TRUE)
```

```
Call: survfit(formula = remsurv ~ remission$treatment)

                       n events *rmean *se(rmean) median 0.95LCL 0.95UCL
remission$treatment=A 26     23   68.5       15.7   31.5      28      74
remission$treatment=B 18     14  109.3       21.8   87.0      20      NA
    * restricted mean with upper limit =  257
```

Now, we'll plot these estimates. The essential conclusion should be that Treatment B shows a generally longer time to remission, making Treatment A appear more attractive.

First, a fairly simple version of this plot.

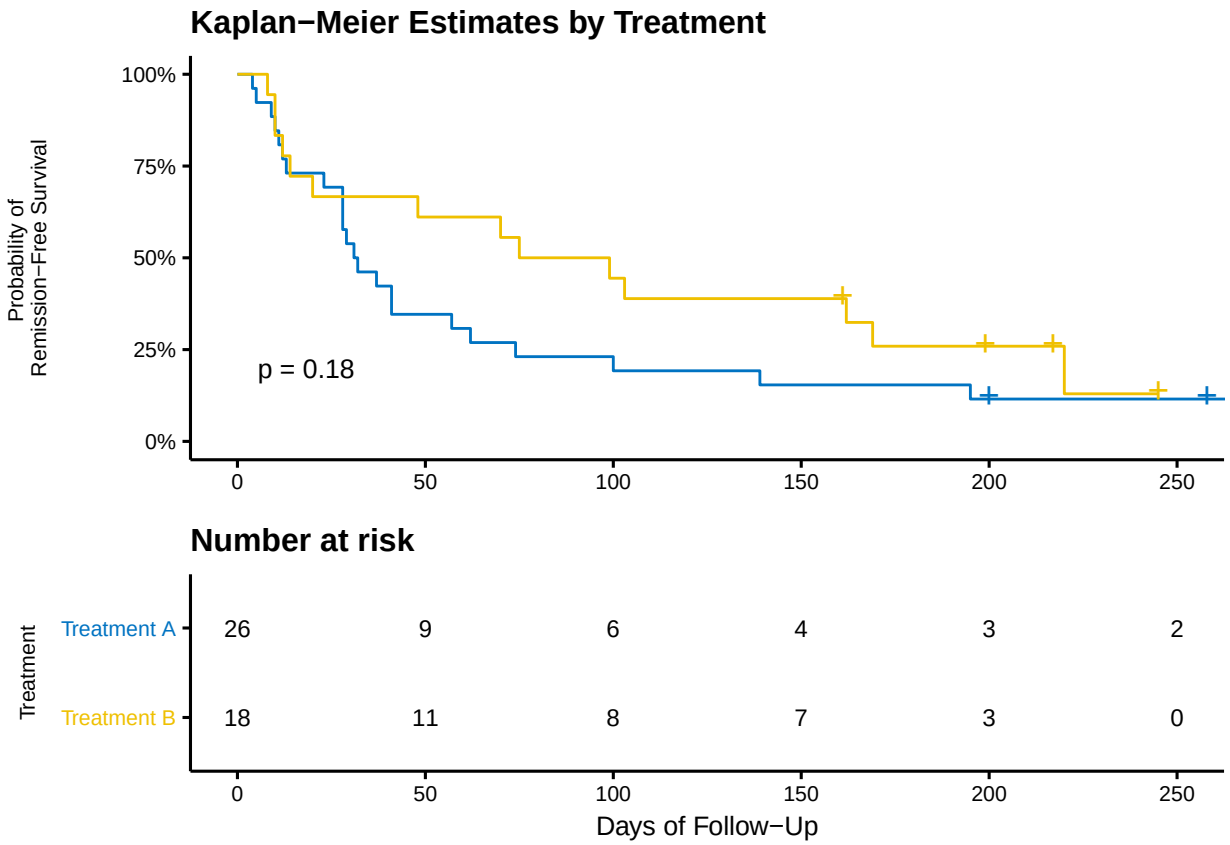
```
ggsurvplot(remfit, data = remission,
           ylab = "Pr(Remission-free survival)",
           xlab = "",
           risk.table = TRUE,
           risk.table.height = 0.25)
```




And now, here's a much more complicated version of the plot.

```
remfit.curve <- ggsurvplot(remfit,
                           data=remission,
                           break.time.by = 50,
                           surv.scale =  "percent",
                           xlab=NULL,
                           ylab="Probability of \nRemission-Free Survival",
                           risk.table = TRUE,
                           legend = "none",
                           legend.title = "Treatment",
```

```
                            size=0.5, conf.int=FALSE,
                            censor = TRUE,
                            palette = "jco",
                            surv.plot.height = 0.5 ,
                            risk.table.height = 0.4,
                            pval = TRUE ,
                            pval.size=3.5,
                            risk.table.fontsize = 3,
                            legend.labs =  c("Treatment A", "Treatment B"))



remfit.curve$plot <- remfit.curve$plot + labs(
  title    = "Kaplan-Meier Estimates by Treatment")

remfit.curve$table <- remfit.curve$table+
 xlab("Days of Follow-Up")+
  labs(title = "Number at risk")

remfit.curve <- ggpar(
  remfit.curve ,
  font.title    = c(12, "bold", "black"),
  font.x        = c(10),
  font.y        = c(8),
  font.xtickslab = c(8, "plain", "black"),
  font.ytickslab = c(8))

remfit.curve
```




## 6.2 Perform and interpret a log rank test, and compare the median and (restricted) mean survival times appropriately.

```
survdiff(remsurv ~ remission$treatment)
```

```
Call:
survdiff(formula = remsurv ~ remission$treatment)

                       N Observed Expected (O-E)^2/E (O-E)^2/V
remission$treatment=A 26       23       19     0.839      1.79
remission$treatment=B 18       14       18     0.886      1.79

 Chisq= 1.8  on 1 degrees of freedom, p= 0.2
```

Treatment A appears to have a better survival profile (shorter time before remission), but the difference between the two treatments' survival functions does not reach the level of statistical significance. This is certainly related to the small sample size in each treatment group.