Biostat 537 HW 2

David Coomes

2/5/2020

Problem 1

(a)

Maintenance Group Estimates

Time (t)	# at risk (n)	# events (d)	d/n	1-d/n	S(t)	H(t)
9	12	1	0.083	0.917	0.917	0.083
12	11	1	0.091	0.909	0.833	0.174
13	10	0	0.000	1.000	0.833	0.174
18	9	1	0.111	0.889	0.741	0.285
23	8	1	0.125	0.875	0.648	0.410
28	7	0	0.000	1.000	0.648	0.410
31	6	1	0.167	0.833	0.540	0.577
34	5	1	0.200	0.800	0.432	0.777
45	4	1	0.250	0.750	0.324	1.027
48	2	1	0.500	0.500	0.162	1.527
161	1	0	0.000	1.000	0.162	1.527

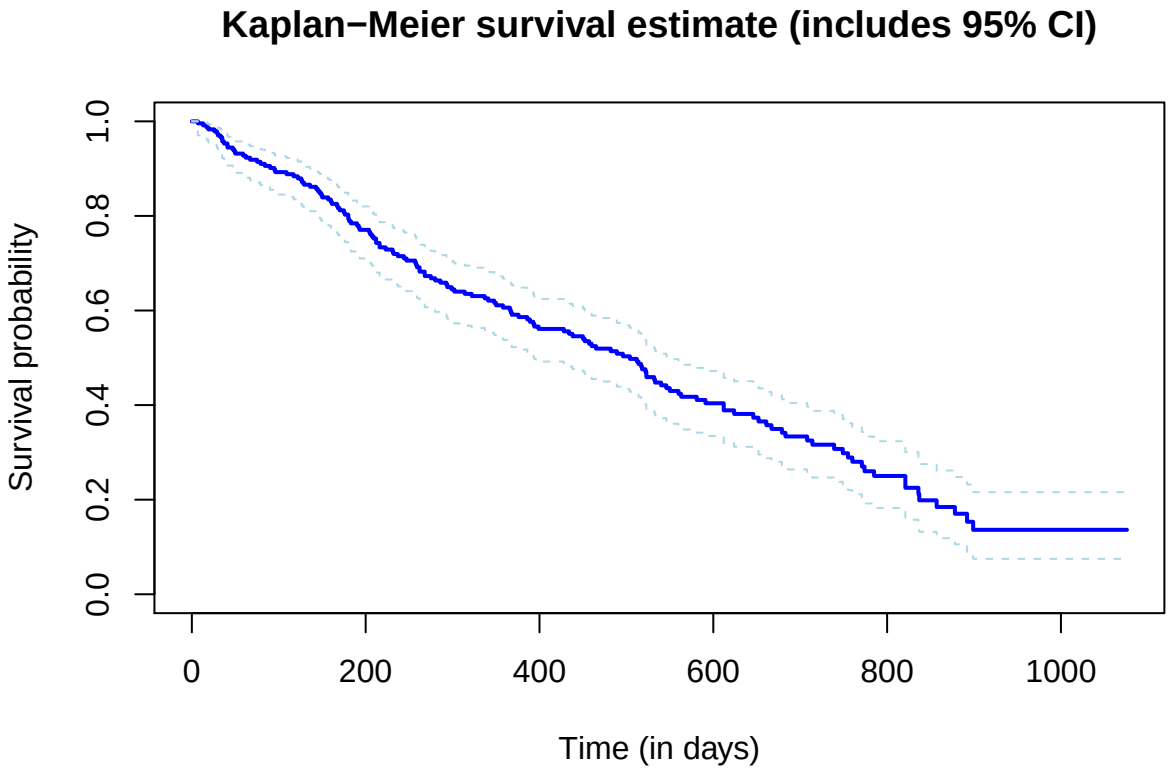
Control Group Estimates

Time (t)	# at risk (n)	# events (d)	d/n	1-d/n	S(t)	H(t)
4	13	1	0.077	0.923	0.923	0.077
5	12	1	0.083	0.917	0.846	0.160
8	11	2	0.182	0.818	0.692	0.342
10	9	0	0.000	1.000	0.692	0.342
12	8	1	0.125	0.875	0.606	0.467
16	7	0	0.000	1.000	0.606	0.467
23	6	1	0.167	0.833	0.505	0.634
27	5	1	0.200	0.800	0.404	0.834
30	4	1	0.250	0.750	0.303	1.084
38	3	1	0.333	0.667	0.202	1.417
43	2	1	0.500	0.500	0.101	1.917
45	1	1	1.000	0.000	0.000	2.917

(b) For the maintenance group, we estimate a 56.8% probability that no relapse will occur by 36 months. For the control group, we estimate a 69.7% probability that no relapse will occur by 36 months.

Problem 2

(a) The estimated probability that no exit will occur by one year is 60.6% (95% CI: 53.8-66.7).



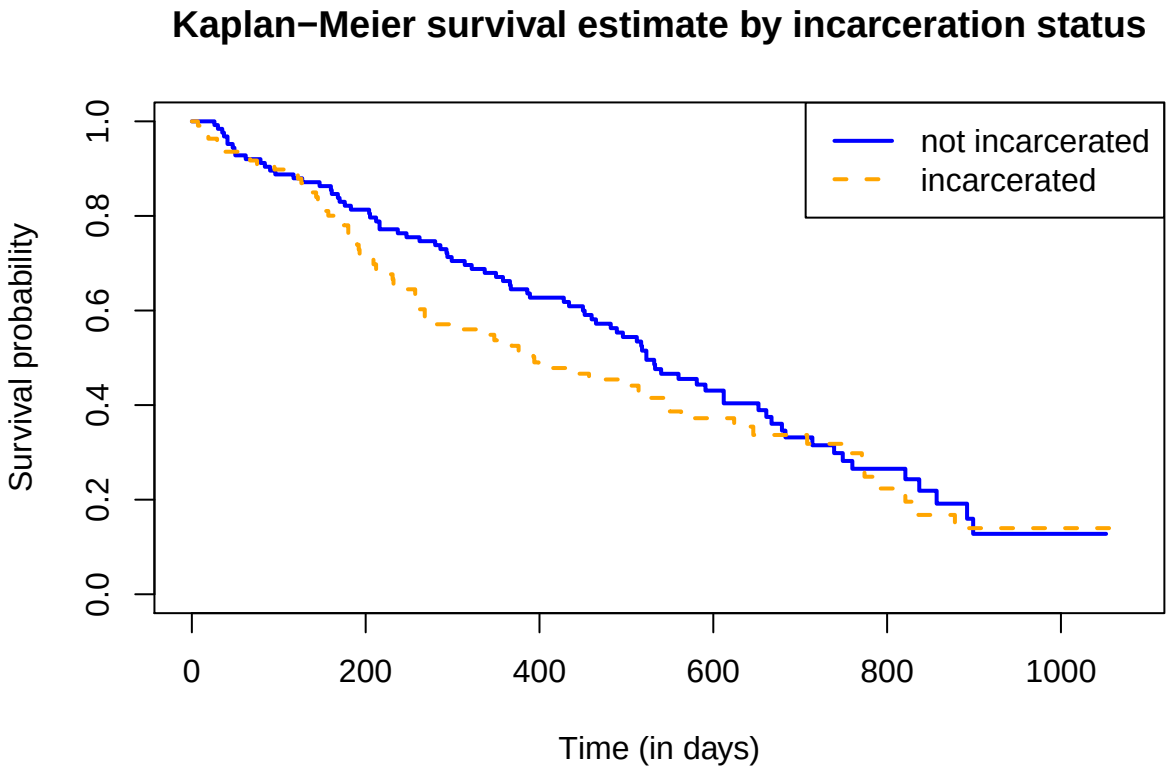
(b) The median time until exit from maintenance is 504 days (95% CI: 394-550).

i. We can obtain the median from looking at the Kaplan-Meier estimator by finding the time where the survival estimate first drops below 0.5. We can estimate the 95% confidence intervals by using the first time period that includes 0.5 in the 95% confidence interval for the lower estimate, and the first time interval that does not include 0.5 that comes after the median time as the upper estimate.

ii. the median estimate and 95% confidence intervals using the *survfit* command returns the same results as the manual examination of the Kaplan-Meier estimators as described in *part i*.

(c)

i.

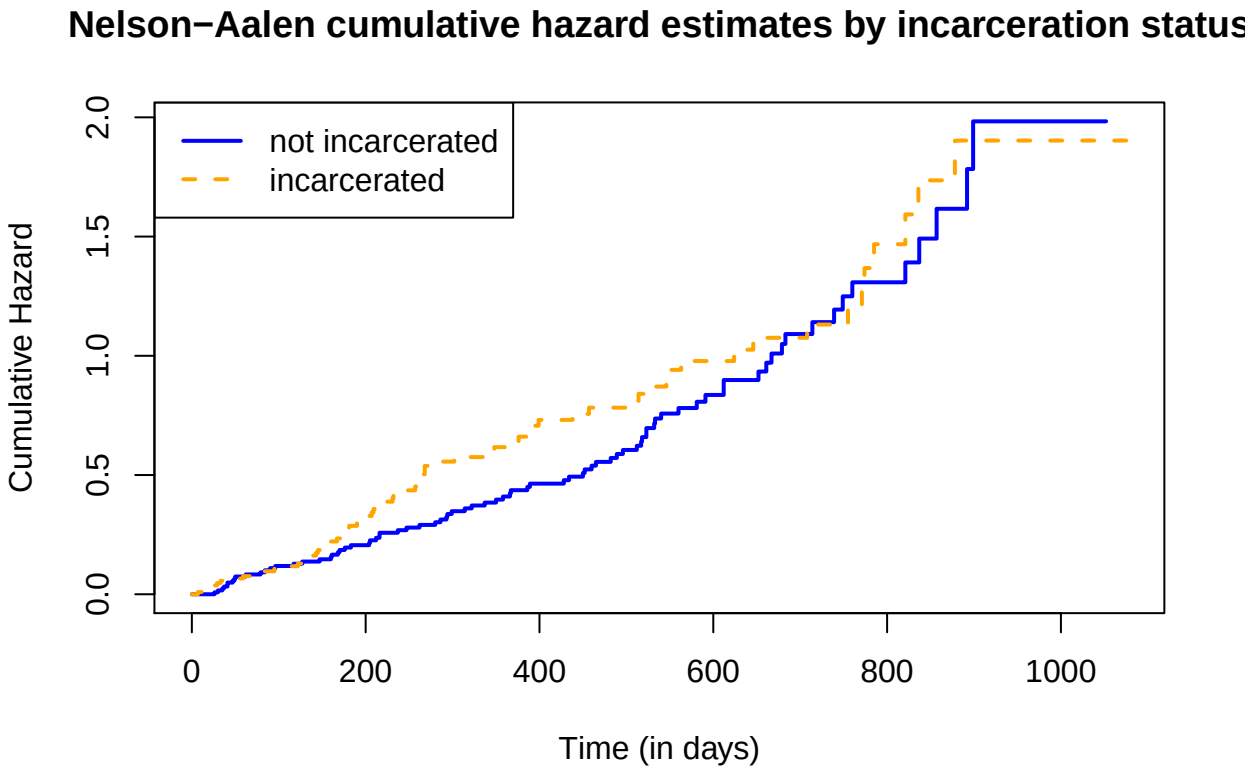


ii. The probability of no exit occuring by 8 months does not significantly differ at the 95% confidence level between those that were incarcerated and those that were not (p=0.077).

iii. Based on the logrank test, the distribution of time until exit from maintenance does not significantly differ by history of incarceration (p=0.3).

iv. Based on the Wilcoxon-Gehan-Breslow test, the survival time until exit from maintenance does not differ significantly by history of incarceration (p=0.11).

v. This plot is informative of the power of the logrank test in that we can see that the two cumulative hazard functions cross over at several points, which reduces the power of the logrank test to detect differences. Since the functions cross over several times towards the end (right side) of the time period, and have some separation towards the beginning and the middle of the time period, I would expect the test statistic for the Wilcoxon-Gehan-Breslow test to be larger than that of the standard logrank test because the former test places more weight on the beginning of the time. We can see that this is true by examining the output of both test statistics from the previous question.



(d)

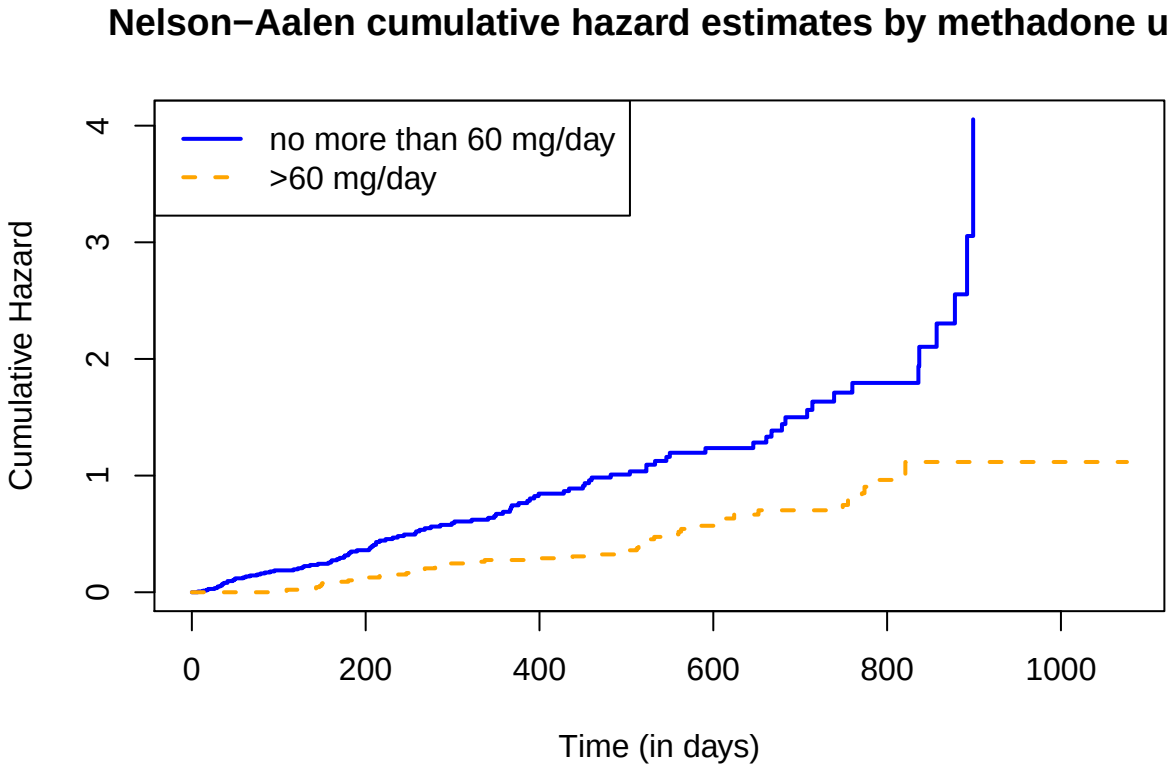
i.

ii. The probability of no exit occuring by 8 months significantly differs at the 95% confidence level between those that had a methadone dose of less than or equal to 60 mg/day and those that recorded a dose of more than 60 mg/day (p=1.57 e-5).

iii. Based on the logrank test, the distribution of time until exit from maintenance significantly differs by methadone use (p=3 e-7).

iv. Based on the Wilcoxon-Gehan-Breslow test, the survival time until exit from maintenance differs significantly by methadone use (p=7.29 e-7).

v. These two cumulative hazard functions do not cross over at any timepoint, therefore we would expect the standard logrank test to have good power in detecting differences between the two groups. I would expect the test statistic for the Wilcoxan-Gehan-Breslow test to be slightly smaller than that of the standard logrank test because there is a larger difference in the two functions towards the later time periods, and the standard logrank test emphasizes this period comparatively. We see that this is true when examining the test statistics in the previous question.



(e) Based on a stratified logrank test, the time until exit from maintenance significantly differs (at the 95% level) based on history of previous incarceration after adjusting for clinic membership (p=0.04).

$$H_0: \; S_{noprison,clinic=1}(t) = S_{prison,clinic=1}(t) \quad AND \quad S_{noprison,clinic=2}(t) = S_{prison,clinic=2}(t) \; for \; all \; t$$

$$H_a: \; At \; least \; one \; of \; the \; above \; equations \; are \; not \; true$$

(f) In calculating the median residual time until exit using only the Kaplan-Meier estimators, I get:

Time (t)	Median residal time
4 months	426 days
8 months	427 days
12 months	389 days

When using the R function I get:

Time (t)	Median residual time
4 months	420 days (95% CI: 376-532)
8 months	427 days (95% CI: 341-520)
12 months	389 days (95% CI: 301-461)

These two are slightly different, and I believe this is because of a small rounding error in which I cannot see

past three digits when looking at the Kaplan-Meier estimators.

Appendix

```
knitr::opts_chunk$set(echo=FALSE, warning=FALSE)

library(pander)
panderOptions('digits', 4)
panderOptions('round', 4)


library(tidyverse)
library(knitr)
library(kableExtra)
source("getmedianres.R")


#Maintenance Group
maint_table_time <- c(9,12,13,18,23,28,31,34,45,48,161)
maint_table_n <- c(12,11,10,9,8,7,6,5,4,2,1)
maint_table_d <- c(1,1,0,1,1,0,1,1,1,1,0)

maint_table <- cbind.data.frame(maint_table_time, maint_table_n, maint_table_d)

maint_table$d_n <- maint_table$maint_table_d / maint_table$maint_table_n
maint_table$d_n_1 <- 1-maint_table$d_n

#Generating survival est
for (val in maint_table$maint_table_time) {
 maint_table <- mutate(maint_table, S_t = ifelse(maint_table_time>9, lag(maint_table$S_t)*d_n_1, 0.9166
}

#Generating cumulative haz est
for (val in maint_table$maint_table_time) {
maint_table <- mutate(maint_table, H_t = ifelse(maint_table_time>9, lag(maint_table$H_t)+d_n, 0.0833333
}

kable(maint_table, digits=3, align="c",
      col.names = c("Time (t)",
                    "# at risk (n)",
                    "# events (d)",
                    "d/n",
                    "1-d/n",
                    "S(t)",
                    "H(t)")) %>%
  kable_styling(bootstrap_options = c("striped", "hover"), full_width = F, position="left") %>%
                column_spec(1, bold=T, border_right=T)


#Control Group
cont_table_time <- c(4,5,8,10,12,16,23,27,30,38,43,45)
cont_table_n <- c(13,12,11,9,8,7,6,5,4,3,2,1)
cont_table_d <- c(1,1,2,0,1,0,1,1,1,1,1,1)
```

```r
cont_table <- cbind.data.frame(cont_table_time, cont_table_n, cont_table_d)

cont_table$d_n <- cont_table$cont_table_d / cont_table$cont_table_n
cont_table$d_n_1 <- 1-cont_table$d_n

#Generating survival est
for (val in cont_table$cont_table_time) {
 cont_table <- mutate(cont_table, S_t = ifelse(cont_table_time>4, lag(cont_table$S_t)*d_n_1, d_n_1))
}

#Generating cumulative haz est
for (val in cont_table$cont_table_time) {
 cont_table <- mutate(cont_table, H_t = ifelse(cont_table_time>4, lag(cont_table$H_t)+d_n, d_n))
}


kable(cont_table, digits=3, align="c",
      col.names = c("Time (t)",
                    "# at risk (n)",
                    "# events (d)",
                    "d/n",
                    "1-d/n",
                    "S(t)",
                    "H(t)")) %>%
  kable_styling(bootstrap_options = c("striped", "hover"), full_width = F, position="left") %>%
                column_spec(1, bold=T, border_right=T)



link = "https://github.com/dmccoomes/Survival/raw/master/Homework%202/addicts.csv"
adix <- read.csv(link)

link = "https://github.com/dmccoomes/Survival/raw/master/Quiz%20section/fitparametric.R"
source(link)



surv.adix <- Surv(time=adix$time, event=adix$event, type="right")
survfit.adix <- survfit(surv.adix ~ 1, data=adix, conf.type="log-log")

summary(survfit.adix)


plot(survfit.adix,
     conf.int=TRUE,
     main="Kaplan-Meier survival estimate (includes 95% CI)",
     ylab="Survival probability", xlab="Time (in days)",
     col=c("blue", "light blue", "light blue"),
     lty=c("solid", "dashed", "dashed"),
     lwd=c(2, 1, 1),
     caption ="caption")
```

```r
summary(survfit.adix)$table


survfit.inc.adix <- survfit(surv.adix ~ prison, data=adix, conf.type="log-log")

plot(survfit.inc.adix,
     conf.int=FALSE,
     main="Kaplan-Meier survival estimate by incarceration status",
     ylab="Survival probability", xlab="Time (in days)",
     col=c("blue", "orange"),
     lty=c("solid", "dashed"),
     lwd=c(2, 2))
legend("topright", c("not incarcerated", "incarcerated"), col=c("blue", "orange"), lwd=c(2,2), lty=c("s


#fit curve for those that were not incarcerated
adix.inc.0 <- adix[adix$prison == 0, ]
surv.adix.noprison <- Surv(time=adix.inc.0$time, event=adix.inc.0$event)
survfit.adix.noprison <- survfit(surv.adix.noprison ~ 1, data=adix.inc.0, conf.type="log-log")

#fit curve for those that were incarcerated
adix.inc.1 <- adix[adix$prison == 1, ]
surv.adix.prison <- Surv(time=adix.inc.1$time, event=adix.inc.1$event)
survfit.adix.prison <- survfit(surv.adix.prison ~ 1, data=adix.inc.1, conf.type="log-log")

#perform wald test for difference
wald.stat.prison.8months <- (summary(survfit.adix.noprison, times=240)["surv"]$surv - summary(survfit.a

wald.stat.prison.8months

#compute wald p-value
2 * pnorm(-abs(wald.stat.prison.8months))



survdiff(surv.adix ~ prison, data=adix, rho=0)


library(survMisc)
comp(ten(survfit.inc.adix))$tests


plot(survfit.inc.adix,
     fun="cumhaz",
     col=c("blue", "orange"), lwd=2, lty=c("solid", "dashed"),
     xlab="Time (in days)", ylab="Cumulative Hazard", main="Nelson-Aalen cumulative hazard estimates by
legend("topleft", c("not incarcerated", "incarcerated"), col=c("blue", "orange"), lwd=c(2,2), lty=c("so


adix$methuse_over60 <- ifelse(adix$dose>60, 1, 0)

survfit.meth.adix <- survfit(surv.adix ~ methuse_over60, data=adix, conf.type="log-log")
```

```r
plot(survfit.meth.adix,
     conf.int=FALSE,
     main="Kaplan-Meier survival estimate by methadone use",
     ylab="Survival probability", xlab="Time (in days)",
     col=c("blue", "orange"),
     lty=c("solid", "dashed"),
     lwd=c(2, 2))
legend("topright", c("no more than 60 mg/day", "> 60 mg/day"), col=c("blue", "orange"), lwd=c(2,2), lty


#fit curve for those that were not incarcerated
adix.meth.0 <- adix[adix$methuse_over60 == 0, ]
surv.adix.meth.0 <- Surv(time=adix.meth.0$time, event=adix.meth.0$event)
survfit.adix.meth.0 <- survfit(surv.adix.meth.0 ~ 1, data=adix.meth.0, conf.type="log-log")

#fit curve for those that were incarcerated
adix.meth.1 <- adix[adix$methuse_over60 == 1, ]
surv.adix.meth.1 <- Surv(time=adix.meth.1$time, event=adix.meth.1$event)
survfit.adix.meth.1 <- survfit(surv.adix.meth.1 ~ 1, data=adix.meth.1, conf.type="log-log")

#perform wald test for difference
wald.stat.meth.8months <- (summary(survfit.adix.meth.0, times=240)["surv"]$surv - summary(survfit.adix.m

wald.stat.meth.8months

#compute wald p-value
2 * pnorm(-abs(wald.stat.meth.8months))



survdiff(surv.adix ~ methuse_over60, data=adix, rho=0)


library(survMisc)
comp(ten(survfit.meth.adix))$tests


plot(survfit.meth.adix,
     fun="cumhaz",
     col=c("blue", "orange"), lwd=2, lty=c("solid", "dashed"),
     xlab="Time (in days)", ylab="Cumulative Hazard", main="Nelson-Aalen cumulative hazard estimates by
legend("topleft", c("no more than 60 mg/day", ">60 mg/day"), col=c("blue", "orange"), lwd=c(2,2), lty=c


survdiff(surv.adix ~ prison + strata(clinic), data=adix)


#Starting with 4 months
summary(survfit.adix, times=120)
0.5 * summary(survfit.adix, times=120)["surv"]$surv
summary(survfit.adix, times=seq(530,560))
546 - 120
#mean residual for 4 months is 426 days
```

```r
#8 months
summary(survfit.adix, times=240)
0.5 * summary(survfit.adix, times=240)["surv"]$surv
summary(survfit.adix, times=seq(600,800))
667-240
# mean residual time for 8 months is 427 days

#12 months
summary(survfit.adix, times=360)
0.5 * summary(survfit.adix, times=360)["surv"]$surv
summary(survfit.adix, times=seq(700, 800))
749-360
#mean residual time for 12 months is 389 days


#4 months
df.for.medresidual <- data.frame("y"=adix$time, "delta"=adix$event)
medresidual.120 <- getmedianres(survobj=surv.adix,
                                times=120, confint=TRUE)

medresidual.120$estimates
c(medresidual.120$ci.lower, medresidual.120$ci.upper)

#8 months
medresidual.240 <- getmedianres(survobj = surv.adix,
                                times=240, confint=TRUE)

medresidual.240$estimates
c(medresidual.240$ci.lower, medresidual.240$ci.upper)

#12 months
medresidual.360 <- getmedianres(survobj = surv.adix,
                                times=360, confint=TRUE)

medresidual.360$estimates
c(medresidual.360$ci.lower, medresidual.360$ci.upper)
```