ST525 HW 4

Nora Quick

Question 1

a) This estimates the probability of observations we KNOW didn't survive the trigger time.

b) This estimates the probability that the observation survives the event time.

c) Yes, the exact values of censoring times plays some role in the K-M estimator.

d) Yes, the exact values of event times plays some role in the K-M estimator because they are "triggers" for the K-M estimator.

e) With no censoring the K-M estimator and survival function simply are the number of observations that survive after t over the sample size.

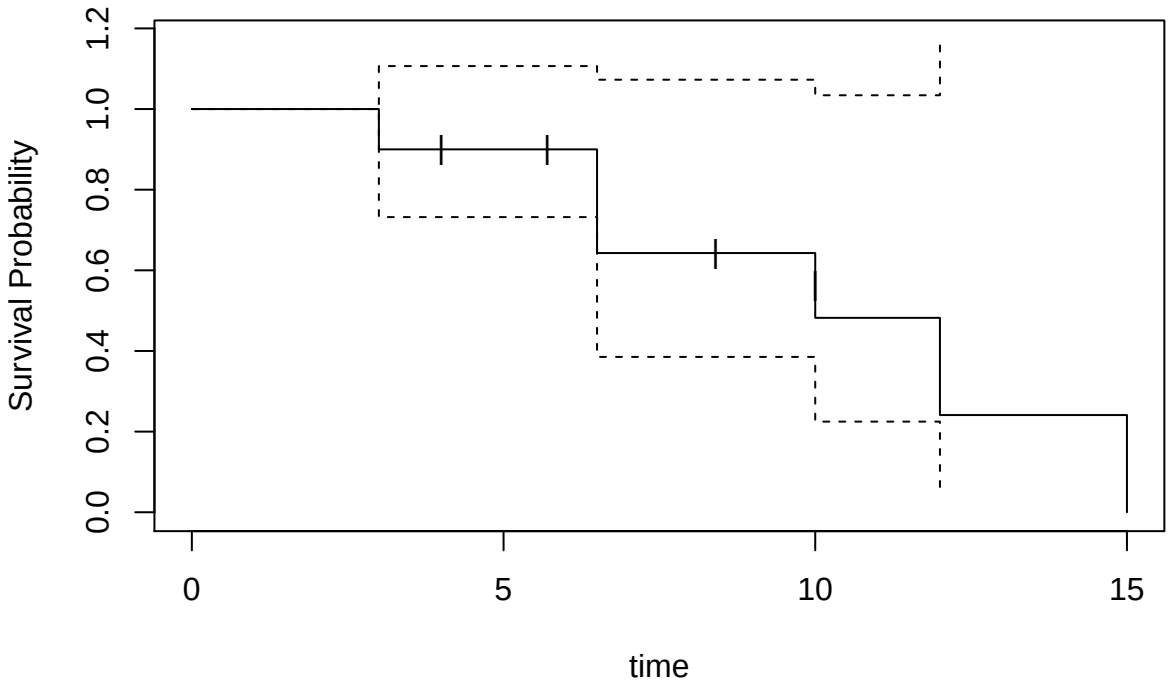
```
obs <- c(1,1,2,3,3,5)
surv1 <- 4/6 # greater than 1
surv2 <- 3/6 # greater than 2
surv3 <- 1/6 # greater than 3
```

Question 2

```
## Loading required package: ggplot2
```

```
## Loading required package: ggpubr
```

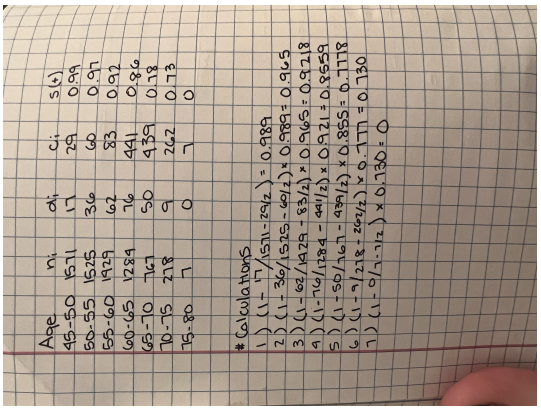
```
##  time  survival   failure Survival.Std.Err No.Left No.Failed No.Censored
##   0.0 1.0000000 0.0000000       0.00000000      10         0           0
##   3.0 0.9000000 0.1000000       0.09486833      10         1           0
##   4.0 0.9000000 0.1000000       0.09486833       9         0           1
##   5.7 0.9000000 0.1000000       0.09486833       8         0           1
##   6.5 0.6428571 0.3571429       0.16794939       7         2           0
##   8.4 0.6428571 0.3571429       0.16794939       5         0           1
##  10.0 0.4821429 0.5178571       0.18771853       4         1           1
##  12.0 0.2410714 0.7589286       0.19459517       2         1           0
##  15.0 0.0000000 1.0000000              NaN       1         1           0
```



b) Based on the above table and looking at the hint provided in the homework that median survival time is 10.0 and the corresponding estimated survival function value is 0.48.

```
## [1] 0.0000000 0.1053605 0.2107210 0.3160815 0.7579143 1.1997471 1.9292619
## [8] 3.3519239       Inf
```

Question 3

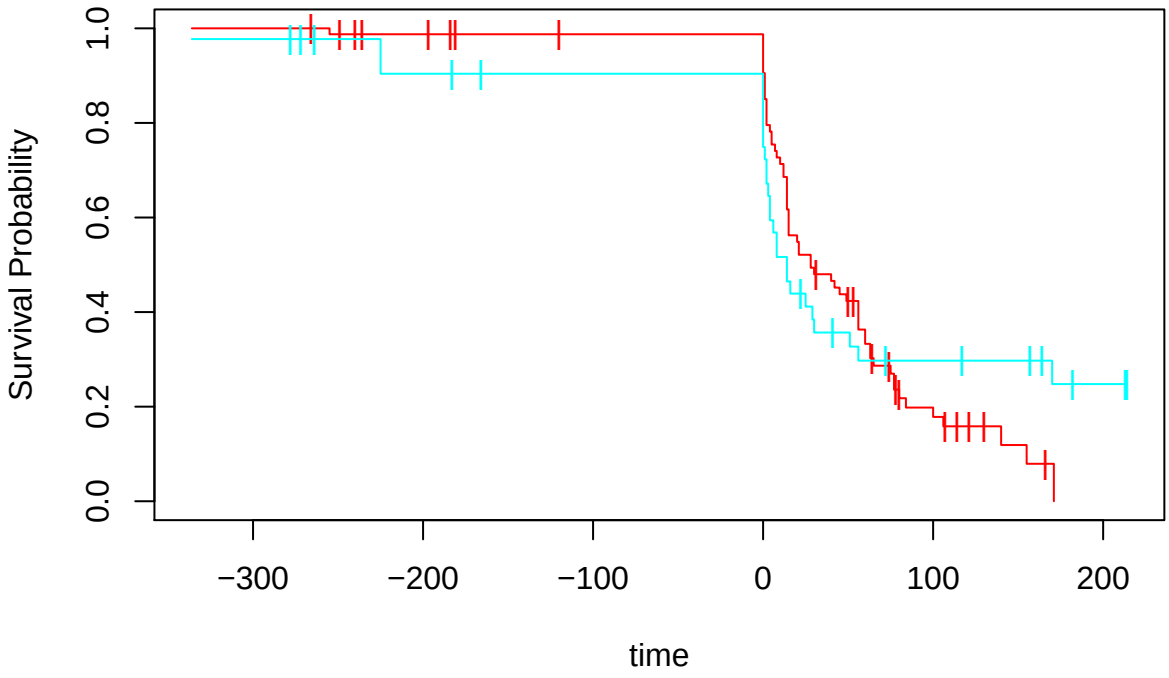


b) The estimated survival probability for interval [55-60] is 0.92.

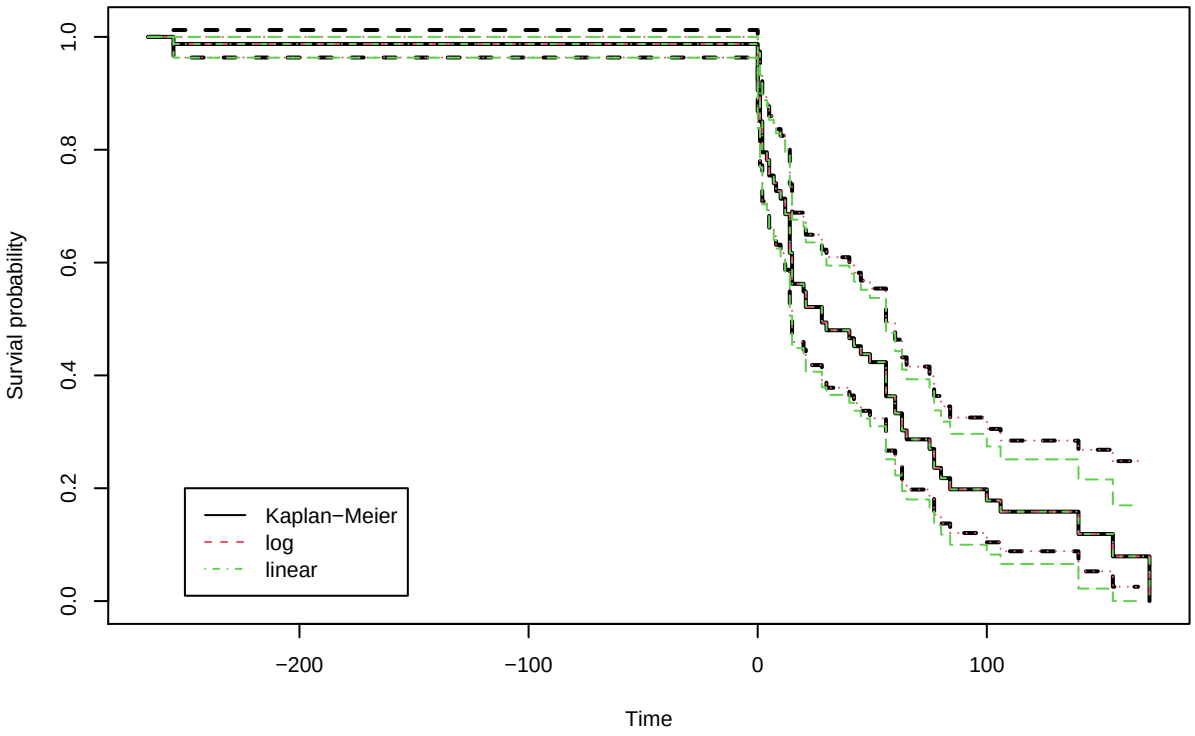
c) My interpritation of this survival probability is that the right before the interval ages of 55-60 there was a 92% probability that the men would not contract CHD.

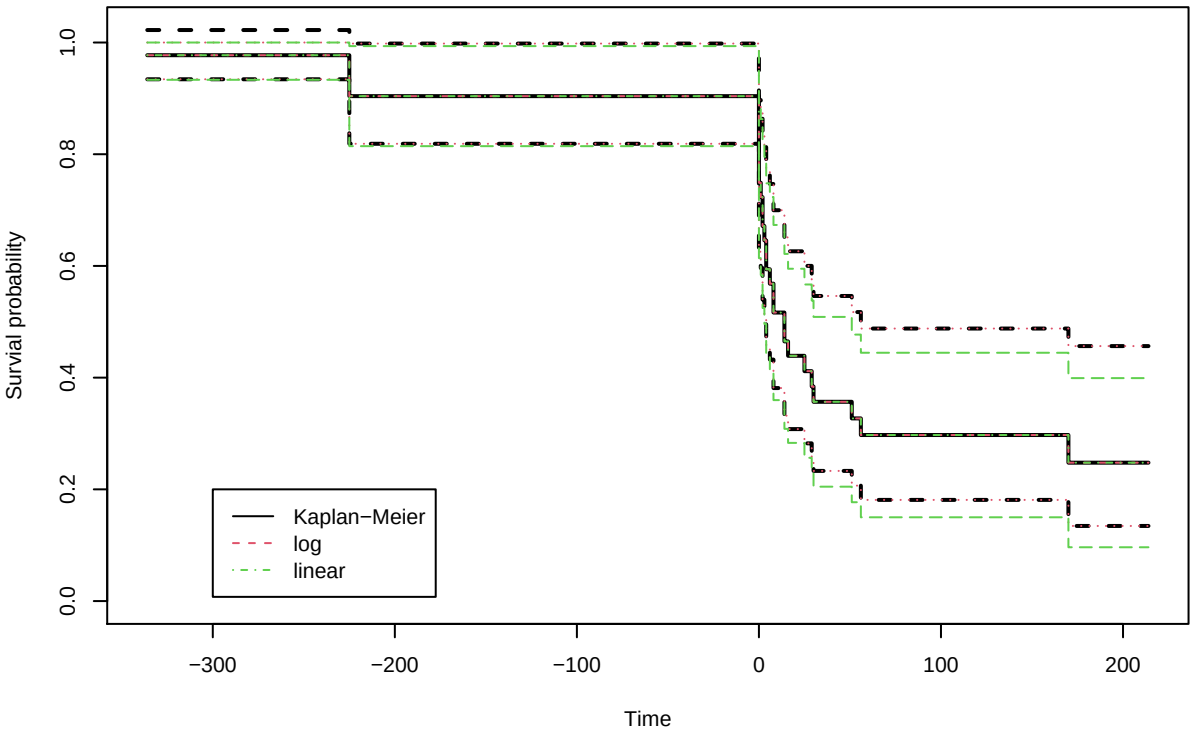
Question 4

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

##    id ttr relapse grp age gender race employment yearsSmoking levelSmoking
## 1  21  41       0   2  36      1    4          1           26             1
## 2 113  14       1   2  41      1    4          2           27             1
## 3  39   5       1   1  25      0    4          2           12             1
## 4  80  16       1   1  54      1    4          1           39             1
## 5  87   0       1   1  45      1    4          2           30             1
## 6  29 157       0   1  43      1    2          1           30             1
##    admitdate      fdate priorAttempts longestNoSmoke time
## 1 2020-11-20 2020-12-31             0              0   41
## 2 2020-06-16 2020-06-30             3             90   14
## 3 2020-05-09 2020-05-14             3             21    5
## 4 2020-10-26 2020-11-11             0              0   16
## 5 2020-09-27 2020-09-27             0              0    0
## 6 2020-07-06 2020-12-10             2           1825  157
```



c) Yes, an exponential distribution would work well for the males (blue like above). While the graph also shows negative times once we hit real times we can see that the male line exponentially decreases.


Survial probability
0.0
0.2
0.4
0.6
0.8
1.0
−200
−100
0
100
Time
Kaplan−Meier
log
linear



e) From the many plots above it appears that the survival experiences of the female, while decreasing more slowly, result in lower survival rates. The males, while having an exponentially decreasing survival rate, the survival rate platoes higher than the females.