```r
if (!require("tidyverse")) install.packages("tidyverse")
```

```
## Loading required package: tidyverse
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.5
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
if (!require("KMsurv")) install.packages("KMsurv")
```

```
## Loading required package: KMsurv
```

```r
if (!require("survival")) install.packages("survival")
```

```
## Loading required package: survival
```

```r
library(KMsurv)
library(survival)
```

```r
data(pneumon)

pneumon <- pneumon %>% arrange(desc(chldage))
head(pneumon)
```

```
##   chldage hospital mthage urban alcohol smoke region poverty bweight race
## 1      12        0     22     1       0     0      1       1       1    1
## 2      12        0     20     1       1     0      1       1       0    1
## 3      12        0     20     1       0     0      1       1       0    1
## 4      12        0     21     1       1     0      1       1       0    1
## 5      12        0     24     1       0     0      1       1       0    1
## 6      12        0     27     1       0     0      1       1       0    1
##   education nsibs wmonth sfmonth agepn
## 1        10     1      1       1     1
## 2        12     1      2       2    12
## 3        12     0      0       0    12
## 4        12     0      3       2    12
## 5        16     0      0       0    12
## 6        16     1      0       0    12
```

#Creating censor variable in pneumon dataset

```r
censor=pneumon$chldage
censor=ifelse(censor >=12,0,1)
table(pneumon$chldage,censor)
```

```
##      censor
##         0   1
##   0.5   0  84
##   1     0 104
##   2     0  98
##   3     0  95
##   4     0  96
##   5     0 113
```

```
##   6      0  101
##   7      0   97
##   8      0   97
##   9      0   89
##   10     0   78
##   11     0   72
##   12  2346    0
```

```
head(pneumon)
```

```
##   chldage hospital mthage urban alcohol smoke region poverty bweight race
## 1      12        0     22     1       0     0      1       1       1    1
## 2      12        0     20     1       1     0      1       1       0    1
## 3      12        0     20     1       0     0      1       1       0    1
## 4      12        0     21     1       1     0      1       1       0    1
## 5      12        0     24     1       0     0      1       1       0    1
## 6      12        0     27     1       0     0      1       1       0    1
##   education nsibs wmonth sfmonth agepn
## 1        10     1      1       1     1
## 2        12     1      2       2    12
## 3        12     0      0       0    12
## 4        12     0      3       2    12
## 5        16     0      0       0    12
## 6        16     1      0       0    12
```

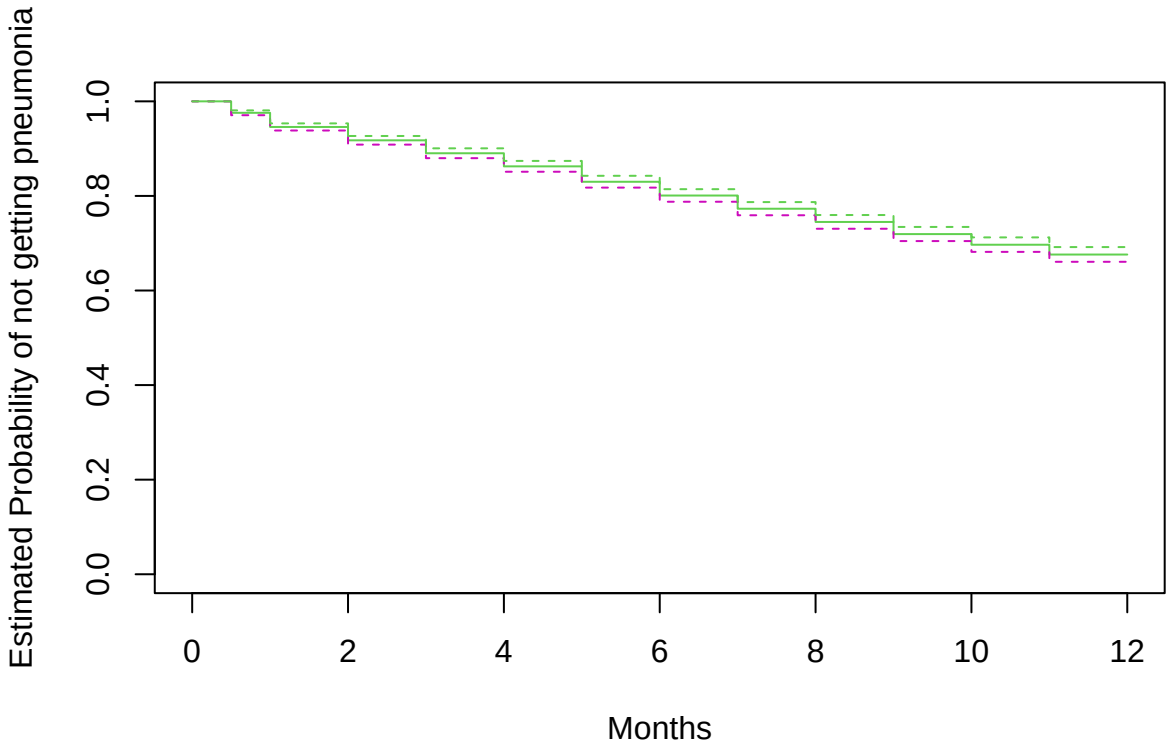
```
pneumon['status'] <- ifelse(pneumon$chldage>=12,0,1)
head(pneumon)
```

```
##   chldage hospital mthage urban alcohol smoke region poverty bweight race
## 1      12        0     22     1       0     0      1       1       1    1
## 2      12        0     20     1       1     0      1       1       0    1
## 3      12        0     20     1       0     0      1       1       0    1
## 4      12        0     21     1       1     0      1       1       0    1
## 5      12        0     24     1       0     0      1       1       0    1
## 6      12        0     27     1       0     0      1       1       0    1
##   education nsibs wmonth sfmonth agepn status
## 1        10     1      1       1     1      0
## 2        12     1      2       2    12      0
## 3        12     0      0       0    12      0
## 4        12     0      3       2    12      0
## 5        16     0      0       0    12      0
## 6        16     1      0       0    12      0
```

Part A

a) Plot the Kaplan-Meier estimator for the survival (not having pneumonia) function for chldage (age at pneumonia).

```
fit.surv <- survfit(Surv(pneumon$chldage, pneumon$status) ~ 1)
plot(fit.surv, xlab = "Months",
    ylab = "Estimated Probability of not getting pneumonia",
    col = c(3,6))
```



#Give an estimate and a confidence interval for a newborn not having developed pneumonia at 6 months.

```
summary(fit.surv, times = c(6))
```

```
## Call: survfit(formula = Surv(pneumon$chldage, pneumon$status) ~ 1)
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##     6   2880     691    0.801 0.00678        0.788        0.814
```

From above summary at 6 months, the estimate of survival would be 0.801 with a confidence interval of 0.788(lower CI)-0.814(upperCI)

Part b

Use the survfit function to stratify survival by poverty (Mother at poverty)

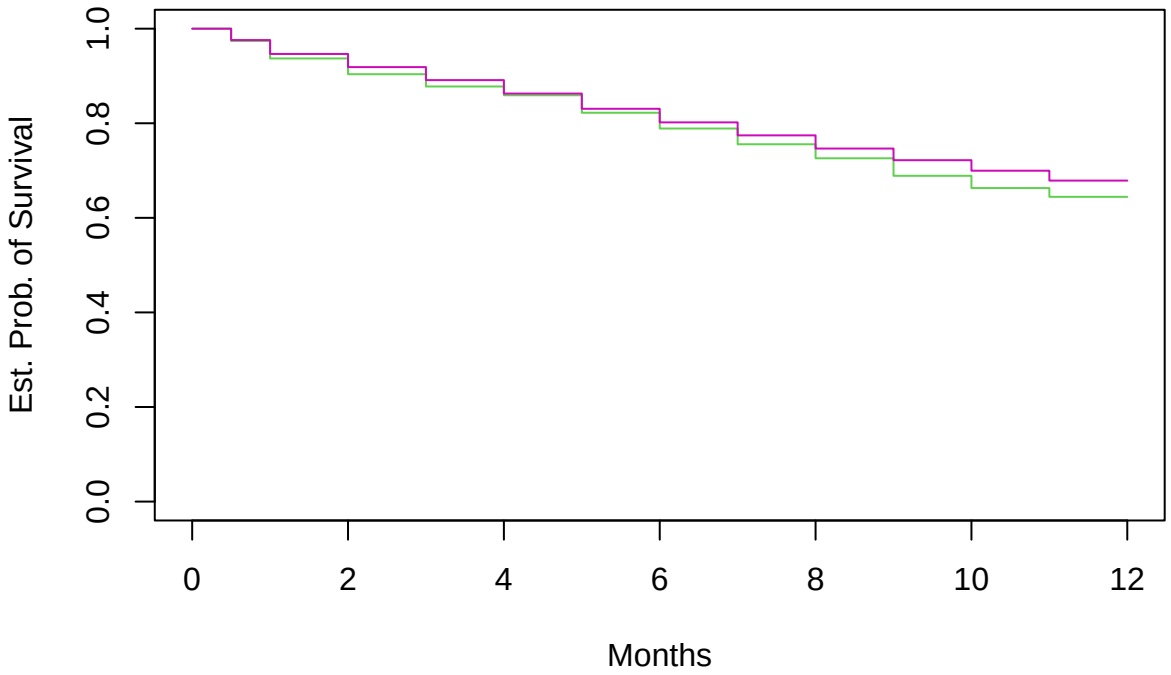
```
fit.pov <- survfit(Surv(pneumon$chldage, pneumon$status) ~ pneumon$poverty)

fit.pov
```

```
## Call: survfit(formula = Surv(pneumon$chldage, pneumon$status) ~ pneumon$poverty)
##
##                       n events median 0.95LCL 0.95UCL
## pneumon$poverty=0   270     96     NA      NA      NA
## pneumon$poverty=1 3200   1028     NA      NA      NA
```

b) Generate the survival curves by poverty (1=yes 0=no) and perform the log-rank test

```
plot (fit.pov , xlab = " Months ",
ylab = " Est. Prob. of Survival ", col = c(3,6))
```




```
logrank.test <- survdiff(Surv(pneumon$chldage, pneumon$status) ~ pneumon$poverty)
logrank.test
```

```
## Call:
## survdiff(formula = Surv(pneumon$chldage, pneumon$status) ~ pneumon$poverty)
##
##                       N Observed Expected (O-E)^2/E (O-E)^2/V
## pneumon$poverty=0  270       96     86.1    1.1476      1.28
## pneumon$poverty=1 3200     1028   1037.9    0.0952      1.28
##
##  Chisq= 1.3  on 1 degrees of freedom, p= 0.3
```

```
#Poverty variable is an int in the dataset but can be used as a factor/categorical value, all calculati

#pneumon$poverty <- as.integer(pneumon$poverty)

#fit.poverty1 <- survfit(Surv(pneumon$chldage, pneumon$status) ~ pneumon$poverty)
#summary(fit.poverty1)
#plot(fit.poverty1, xlab = "Months",
#     ylab = "Estimated Probability of not getting pneumonia using poverty as int", col = c(3,6))

#pneumon$poverty <- as.factor(pneumon$poverty)

#fit.poverty2 <- survfit(Surv(pneumon$chldage, pneumon$status) ~ pneumon$poverty)
#summary(fit.poverty2)
#plot(fit.poverty2, xlab = "Months",
#    ylab = "Estimated Probability of not getting pneumonia using poverty as factors", col = c(3,6))
```

#part c #Fit the Cox proportional hazards model that includes only poverty as a covariate use ties='breslow' option

```
#Fit the Cox proportional hazards model that includes only poverty as a covariate use ties='breslow' op

#without ties
```

```r
fit.coxbasic <- coxph(Surv(pneumon$chldage, pneumon$status) ~ pneumon$poverty)
summary(fit.coxbasic)
```

```
## Call:
## coxph(formula = Surv(pneumon$chldage, pneumon$status) ~ pneumon$poverty)
##
##   n= 3470, number of events= 1124
##
##                     coef exp(coef) se(coef)      z Pr(>|z|)
## pneumon$poverty -0.1211    0.8860   0.1067 -1.134    0.257
##
##                 exp(coef) exp(-coef) lower .95 upper .95
## pneumon$poverty     0.886      1.129    0.7188     1.092
##
## Concordance= 0.504  (se = 0.004 )
## Likelihood ratio test= 1.24  on 1 df,   p=0.3
## Wald test            = 1.29  on 1 df,   p=0.3
## Score (logrank) test = 1.29  on 1 df,   p=0.3
```

```r
fit.coxbasic
```

```
## Call:
## coxph(formula = Surv(pneumon$chldage, pneumon$status) ~ pneumon$poverty)
##
##                     coef exp(coef) se(coef)      z     p
## pneumon$poverty -0.1211    0.8860   0.1067 -1.134 0.257
##
## Likelihood ratio test=1.24  on 1 df, p=0.2646
## n= 3470, number of events= 1124
```

#Is there a difference in survival between poverty levels (yes or no)

```r
#with ties

fit.coxties <- coxph(Surv(pneumon$chldage, pneumon$status) ~ pneumon$poverty, ties = "breslow")
summary(fit.coxties)
```

```
## Call:
## coxph(formula = Surv(pneumon$chldage, pneumon$status) ~ pneumon$poverty,
##     ties = "breslow")
##
##   n= 3470, number of events= 1124
##
##                     coef exp(coef) se(coef)      z Pr(>|z|)
## pneumon$poverty -0.1189    0.8879   0.1067 -1.114    0.265
##
##                 exp(coef) exp(-coef) lower .95 upper .95
## pneumon$poverty    0.8879      1.126    0.7203     1.094
##
## Concordance= 0.504  (se = 0.004 )
## Likelihood ratio test= 1.2  on 1 df,   p=0.3
## Wald test            = 1.24  on 1 df,   p=0.3
## Score (logrank) test = 1.24  on 1 df,   p=0.3
```

```r
fit.coxties
```

```
## Call:
```

```
## coxph(formula = Surv(pneumon$chldage, pneumon$status) ~ pneumon$poverty,
##     ties = "breslow")
##
##                     coef exp(coef) se(coef)      z     p
## pneumon$poverty -0.1189    0.8879   0.1067 -1.114 0.265
##
## Likelihood ratio test=1.2  on 1 df, p=0.2731
## n= 3470, number of events= 1124
```

No, there is no difference hence no clear evidence for a difference in survival due to pneumonia between poverty groups with even using ties.

```
#Fit a Cox proportional hazards model that includes
# mthage, urban, alcohol, smoke, region, poverty, bweight, race, and education as predictors. use ties=
fit.coxmultiplepred <- coxph(Surv(chldage, status) ~ mthage + urban + alcohol + smoke + region + poverty
summary(fit.coxmultiplepred)
```

```
## Call:
## coxph(formula = Surv(chldage, status) ~ mthage + urban + alcohol +
##     smoke + region + poverty + bweight + race + education, data = pneumon,
##     ties = "breslow")
##
##   n= 3470, number of events= 1124
##
##                coef exp(coef)  se(coef)      z Pr(>|z|)
## mthage     0.164735  1.179081  0.011939 13.798   <2e-16 ***
## urban     -0.096951  0.907601  0.071454 -1.357   0.1748
## alcohol   -0.037410  0.963281  0.028806 -1.299   0.1940
## smoke      0.095499  1.100208  0.047381  2.016   0.0438 *
## region    -0.007363  0.992664  0.031183 -0.236   0.8133
## poverty   -0.147013  0.863283  0.107653 -1.366   0.1721
## bweight    0.110808  1.117180  0.067310  1.646   0.0997 .
## race       0.064341  1.066456  0.043664  1.474   0.1406
## education  0.002859  1.002863  0.016512  0.173   0.8625
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##           exp(coef) exp(-coef) lower .95 upper .95
## mthage       1.1791     0.8481    1.1518     1.207
## urban        0.9076     1.1018    0.7890     1.044
## alcohol      0.9633     1.0381    0.9104     1.019
## smoke        1.1002     0.9089    1.0026     1.207
## region       0.9927     1.0074    0.9338     1.055
## poverty      0.8633     1.1584    0.6991     1.066
## bweight      1.1172     0.8951    0.9791     1.275
## race         1.0665     0.9377    0.9790     1.162
## education    1.0029     0.9971    0.9709     1.036
##
## Concordance= 0.625  (se = 0.008 )
## Likelihood ratio test= 217.8  on 9 df,   p=<2e-16
## Wald test            = 217.5  on 9 df,   p=<2e-16
## Score (logrank) test = 221.2  on 9 df,   p=<2e-16
```

```
fit.coxmultiplepred
```

```
## Call:
```

```
## coxph(formula = Surv(chldage, status) ~ mthage + urban + alcohol +
##     smoke + region + poverty + bweight + race + education, data = pneumon,
##     ties = "breslow")
##
##                coef exp(coef)  se(coef)      z      p
## mthage     0.164735  1.179081  0.011939 13.798 <2e-16
## urban     -0.096951  0.907601  0.071454 -1.357 0.1748
## alcohol   -0.037410  0.963281  0.028806 -1.299 0.1940
## smoke      0.095499  1.100208  0.047381  2.016 0.0438
## region    -0.007363  0.992664  0.031183 -0.236 0.8133
## poverty   -0.147013  0.863283  0.107653 -1.366 0.1721
## bweight    0.110808  1.117180  0.067310  1.646 0.0997
## race       0.064341  1.066456  0.043664  1.474 0.1406
## education  0.002859  1.002863  0.016512  0.173 0.8625
##
## Likelihood ratio test=217.8  on 9 df, p=< 2.2e-16
## n= 3470, number of events= 1124
```

```
# List the significant predictors.
```

Ans The significant predictors are mother age, smoke