Introduction to Survival Analysis

Final Project

Darren Kang Wan Chee - 5003201184

Abima Aunur Rochman - 5003201035

Define Packages

```
library(KMsurv)
library(dplyr)
```

```
##
## 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
```

```
library(survival)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v forcats   1.0.0     v readr     2.1.4
## v ggplot2   3.4.3     v stringr   1.5.0
## v lubridate 1.9.3     v tibble    3.2.1
## v purrr     1.0.2     v tidyr     1.3.0
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

Descriptive Statistics

```
data('pneumon')
pneumon %>% summarise(across(where(is.numeric), .fns =
                                  list(min = min,
                                       median = median,
                                       mean = mean,
                                       stdev = sd,
                                       q25 = ~quantile(., 0.25),
                                       q75 = ~quantile(., 0.75),
                                       max = max))) %>%
  pivot_longer(everything(), names_sep='_', names_to=c('variable', '.value'))
```

```
## # A tibble: 15 x 8
##    variable    min median    mean stdev   q25   q75   max
##    <chr>     <dbl>  <dbl>   <dbl> <dbl> <dbl> <dbl> <dbl>
##  1 chldage     0.5     12  9.84   3.62      8    12    12
##  2 hospital    0        0  0.0210 0.144     0     0     1
##  3 mthage     14       22 21.6    2.72     20    23    29
##  4 urban       0        1  0.761  0.427     1     1     1
##  5 alcohol     0        0  0.665  1.08      0     1     4
##  6 smoke       0        0  0.441  0.668     0     1     2
##  7 region      1        3  2.65   0.961     2     3     4
##  8 poverty     0        1  0.922  0.268     1     1     1
##  9 bweight     0        0  0.360  0.480     0     1     1
## 10 race        1        1  1.61   0.757     1     2     3
## 11 education   0       12 11.4    2.00     10    12    19
## 12 nsibs       0        0  0.678  0.859     0     1     6
## 13 wmonth      0        0  1.93   3.64      0     2    28
## 14 sfmonth     0        0  1.12   1.99      0     1    18
## 15 agepn       0       10  7.86   4.46      3    12    12
```

Data Preprocessing & Cleaning

```
str(pneumon)
```

```
## 'data.frame':    3470 obs. of  15 variables:
##  $ chldage  : num  12 12 3 2 4 12 7 3 7 12 ...
##  $ hospital : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ mthage   : int  22 20 24 22 21 20 24 24 26 21 ...
##  $ urban    : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ alcohol  : int  0 1 3 2 1 0 0 3 2 1 ...
##  $ smoke    : int  0 0 0 2 2 0 0 0 2 0 ...
##  $ region   : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ poverty  : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ bweight  : int  1 0 0 0 1 0 0 0 0 0 ...
##  $ race     : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ education: int  10 12 12 9 12 12 12 14 12 12 ...
##  $ nsibs    : int  1 1 2 0 0 0 1 0 0 0 ...
##  $ wmonth   : int  1 2 1 0 0 0 0 4 1 3 ...
##  $ sfmonth  : int  1 2 0 0 0 0 0 2 1 2 ...
##  $ agepn    : int  1 12 3 2 4 12 7 3 6 12 ...
```

```
sapply(pneumon, function(x) sum(is.na(x)))
```

```
##   chldage  hospital    mthage     urban   alcohol     smoke    region   poverty
##         0         0         0         0         0         0         0         0
##   bweight      race education     nsibs    wmonth   sfmonth     agepn
##         0         0         0         0         0         0         0
```

```
pneumon$mthage <- as.numeric(pneumon$mthage)
pneumon$urban <- as.factor(pneumon$urban)
pneumon$alcohol <- as.factor(pneumon$alcohol)
pneumon$smoke <- as.factor(pneumon$smoke)
pneumon$region <- as.factor(pneumon$region)
```

```r
pneumon$poverty <- as.factor(pneumon$poverty)
pneumon$bweight <- as.factor(pneumon$bweight)
pneumon$race <- as.factor(pneumon$race)
pneumon$education <- as.numeric(pneumon$education)
pneumon$nsibs <- as.numeric(pneumon$nsibs)
pneumon$wmonth <- as.numeric(pneumon$wmonth)
pneumon$sfmonth <- as.numeric(pneumon$sfmonth)
pneumon$smoke <- factor(pneumon$smoke, levels = c(0,1,2), labels = c(0,1,1))
# smoke = 0: no, 1:yes
pneumon$alcohol <- factor(pneumon$alcohol, levels = c(0,1,2,3,4), labels = c(0,1,1,1,1))
# alcohol = 0: no, 1:yes
pneumon$education_cat <- cut(pneumon$education, breaks = c(0,10,12,20), labels =c("Low_Edu", "Med_Edu",
pneumon$wmonth_cat <- cut(pneumon$wmonth, breaks = c(-1,0,30), labels = c("zero_month", "morethan_zero"
pneumon$sfmonth_cat <- cut(pneumon$sfmonth, breaks = c(-1,0,30), labels = c("zero_month", "morethan_zer
remove_cols <- c('education','wmonth','sfmonth')
pneumon = subset(pneumon, select = !(names(pneumon) %in% remove_cols))
str(pneumon)
```

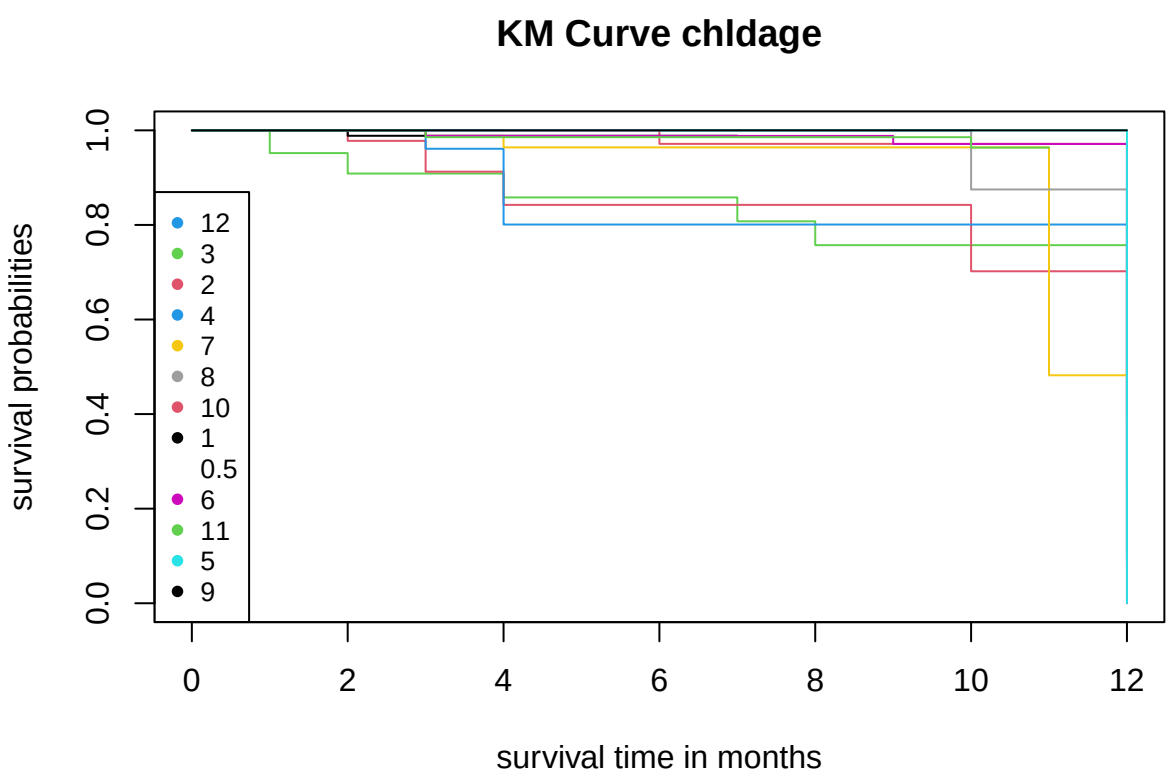
```
## 'data.frame':    3470 obs. of  15 variables:
##  $ chldage      : num  12 12 3 2 4 12 7 3 7 12 ...
##  $ hospital     : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ mthage       : num  22 20 24 22 21 20 24 24 26 21 ...
##  $ urban        : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
##  $ alcohol      : Factor w/ 2 levels "0","1": 1 2 2 2 2 1 1 2 2 2 ...
##  $ smoke        : Factor w/ 2 levels "0","1": 1 1 1 2 2 1 1 1 2 1 ...
##  $ region       : Factor w/ 4 levels "1","2","3","4": 1 1 1 1 1 1 1 1 1 1 ...
##  $ poverty      : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
##  $ bweight      : Factor w/ 2 levels "0","1": 2 1 1 1 2 1 1 1 1 1 ...
##  $ race         : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1 1 1 ...
##  $ nsibs        : num  1 1 2 0 0 0 1 0 0 0 ...
##  $ agepn        : int  1 12 3 2 4 12 7 3 6 12 ...
##  $ education_cat: Factor w/ 3 levels "Low_Edu","Med_Edu",..: 1 2 2 1 2 2 2 3 2 2 ...
##  $ wmonth_cat   : Factor w/ 2 levels "zero_month","morethan_zero": 2 2 2 1 1 1 1 2 2 2 ...
##  $ sfmonth_cat  : Factor w/ 2 levels "zero_month","morethan_zero": 2 2 1 1 1 1 1 2 2 2 ...
```

Define Survival Time

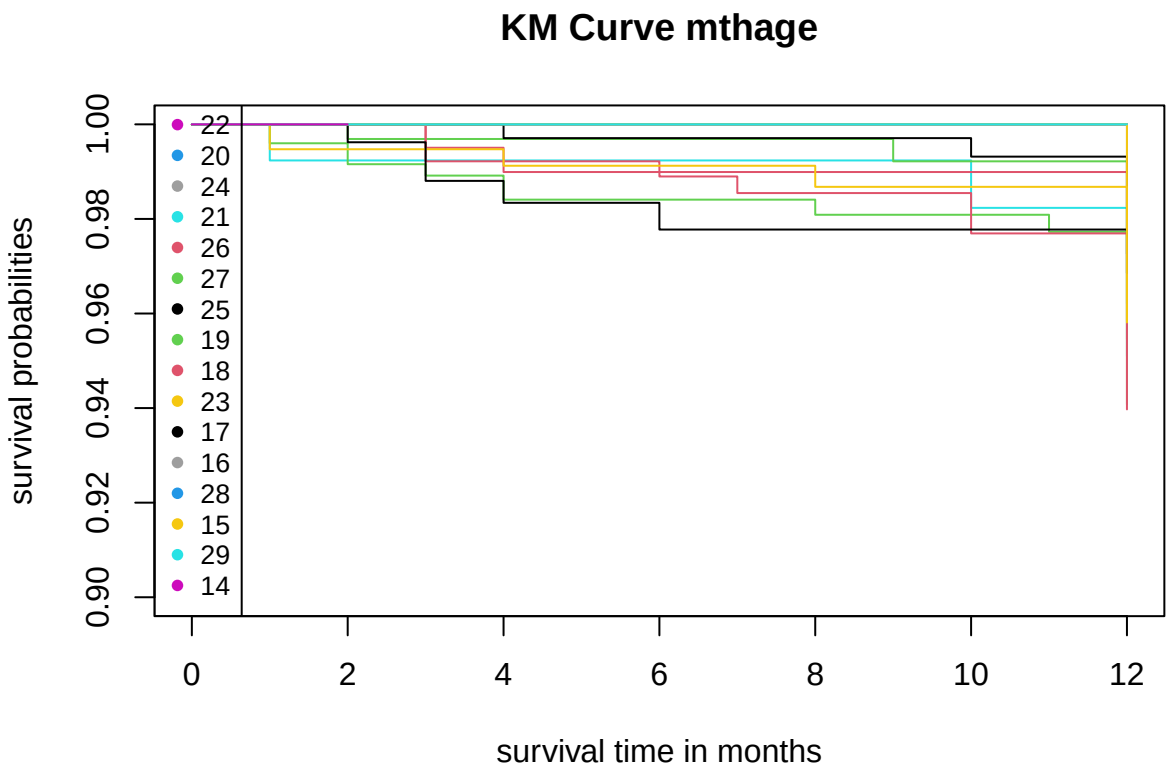
```r
y <- Surv(pneumon$agepn, pneumon$hospital)
```

Plot KM Curve

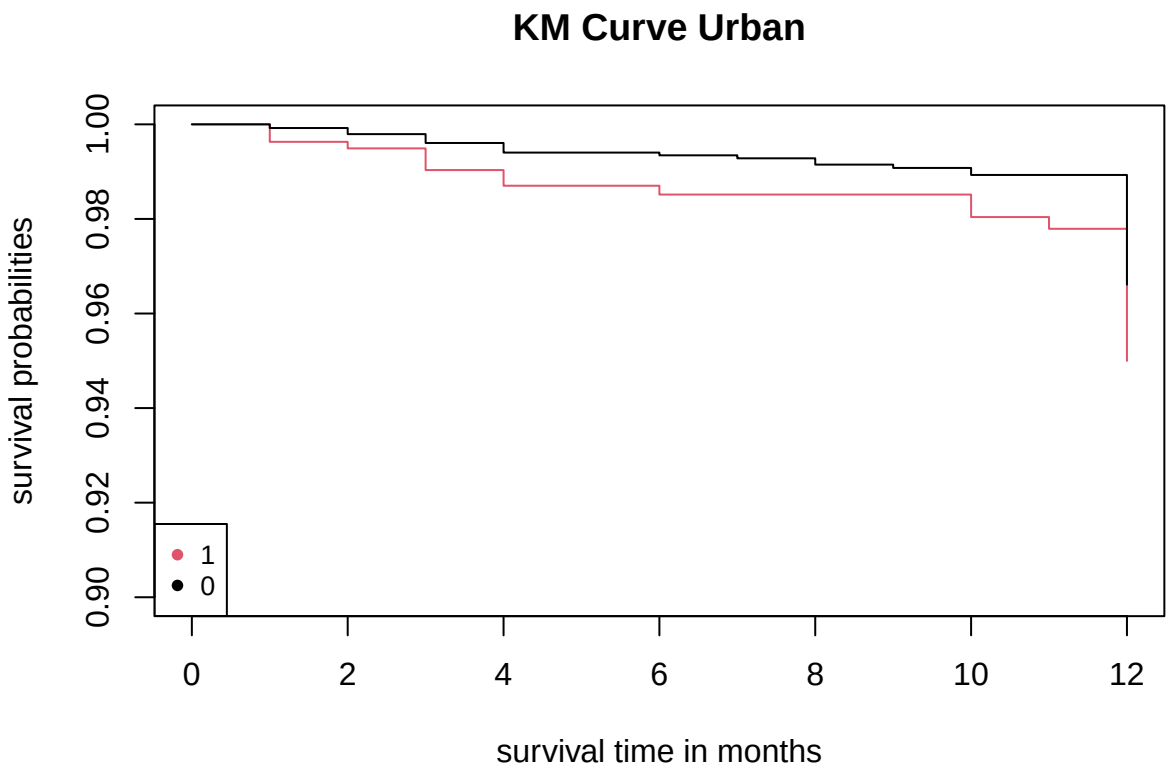
```r
# child age
kmfit1 <- survfit(y ~ chldage, data = pneumon)
plot(kmfit1, col = c(unique(pneumon$chldage)),xlab="survival time in months", ylab="survival probabilit
title(main = "KM Curve chldage")
legend("bottomleft", legend = c(unique(pneumon$chldage)), col=c(unique(pneumon$chldage)), pch = 16,
       cex = 0.8, xjust = 1)
```



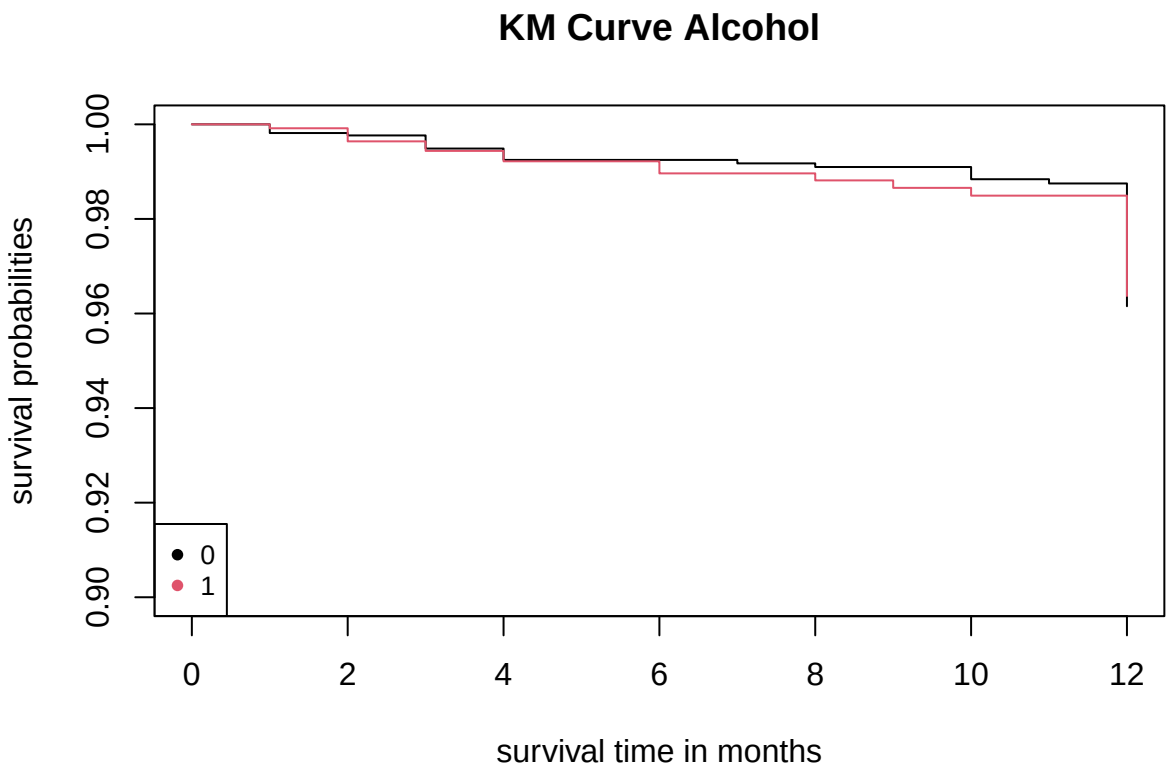

```
# mthage
kmfit2 <- survfit(y ~ mthage, data = pneumon)
plot(kmfit2, col = c(unique(pneumon$mthage)),xlab="survival time in months", ylab="survival probabiliti
     ylim = c(0.9,1))
title(main = "KM Curve mthage")
legend("bottomleft", legend = c(unique(pneumon$mthage)), col=c(unique(pneumon$mthage)), pch = 16,
       cex = 0.8, xjust = 1)
```



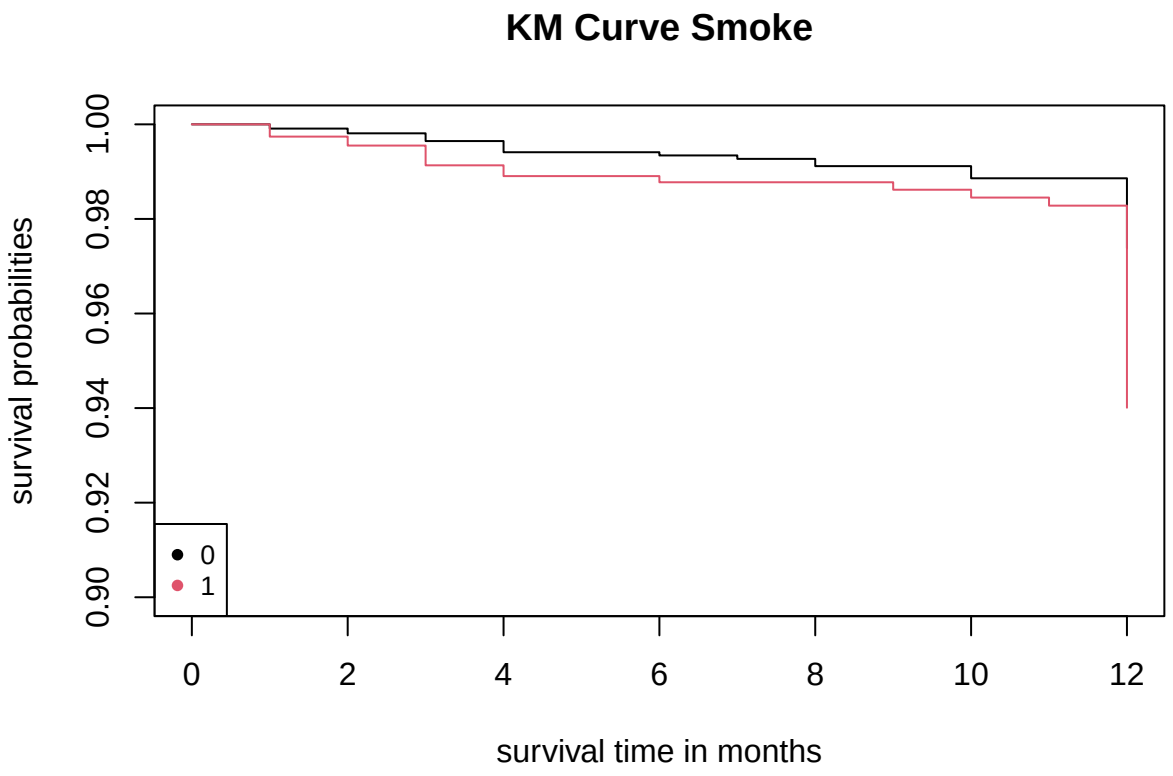

```
# urban
kmfit3 <- survfit(y ~ urban, data = pneumon)
plot(kmfit3, col = c(unique(pneumon$urban)),xlab="survival time in months", ylab="survival probabilities
     ylim = c(0.9,1))
title(main = "KM Curve Urban")
legend("bottomleft", legend = c(unique(pneumon$urban)), col=c(unique(pneumon$urban)), pch = 16,
       cex = 0.8, xjust = 1)
```



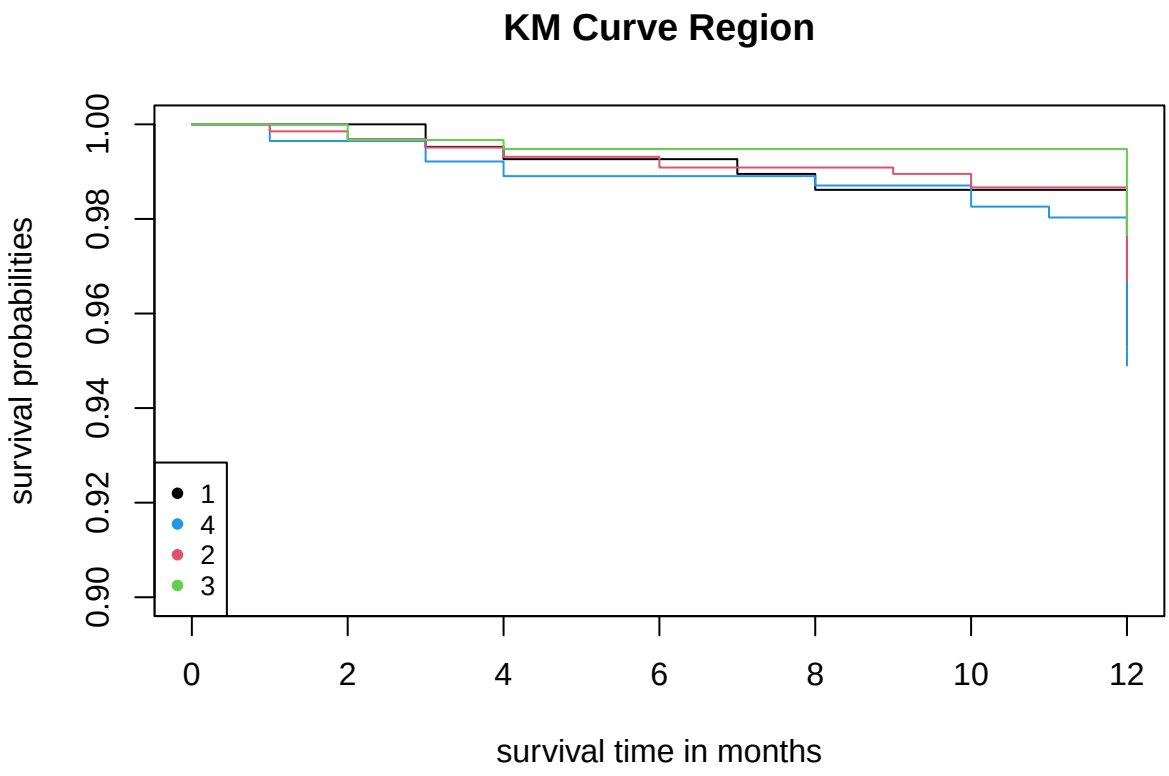

```
# alchohol
kmfit4 <- survfit(y ~ alcohol, data = pneumon)
plot(kmfit4, col = c(unique(pneumon$alcohol)),xlab="survival time in months", ylab="survival probabilit
     ylim = c(0.9,1))
title(main = "KM Curve Alcohol")
legend("bottomleft", legend = c(unique(pneumon$alcohol)), col=c(unique(pneumon$alcohol)), pch = 16,
       cex = 0.8, xjust = 1)
```



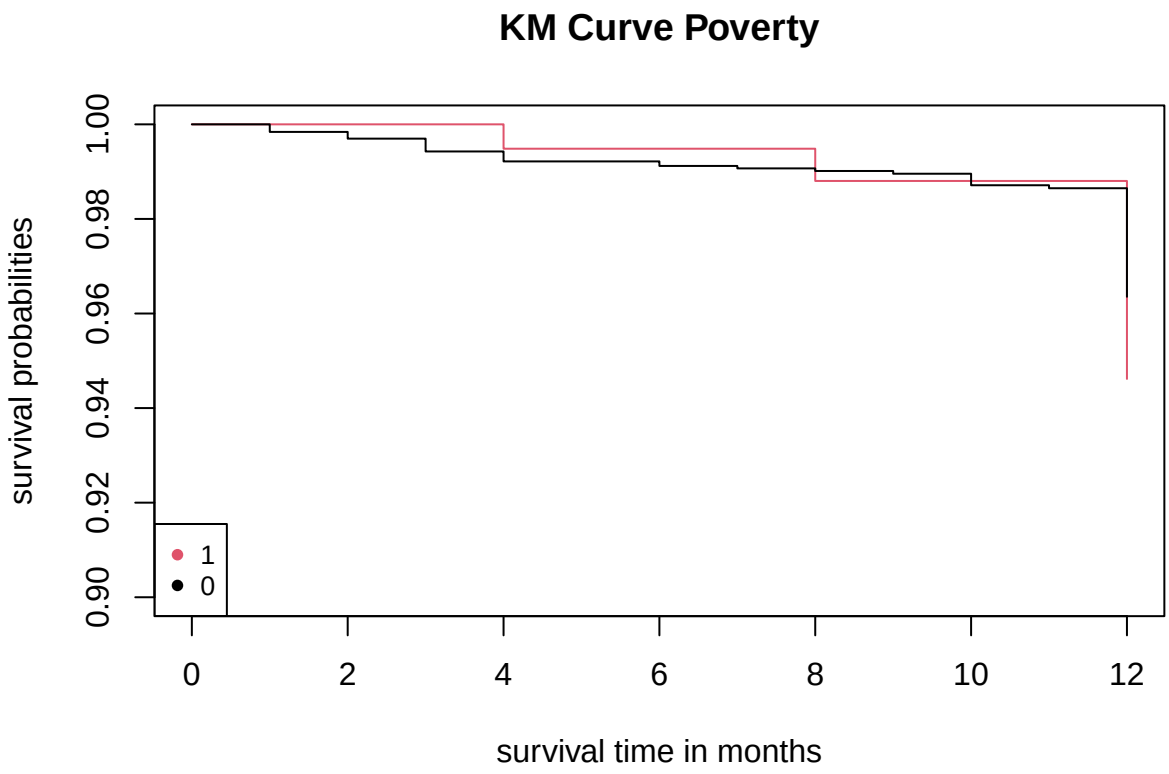

```
# smoke
kmfit5 <- survfit(y ~ smoke, data = pneumon)
plot(kmfit5, col = c(unique(pneumon$smoke)),xlab="survival time in months", ylab="survival probabilities
     ylim = c(0.9,1))
title(main = "KM Curve Smoke")
legend("bottomleft", legend = c(unique(pneumon$smoke)), col=c(unique(pneumon$smoke)), pch = 16,
       cex = 0.8, xjust = 1)
```



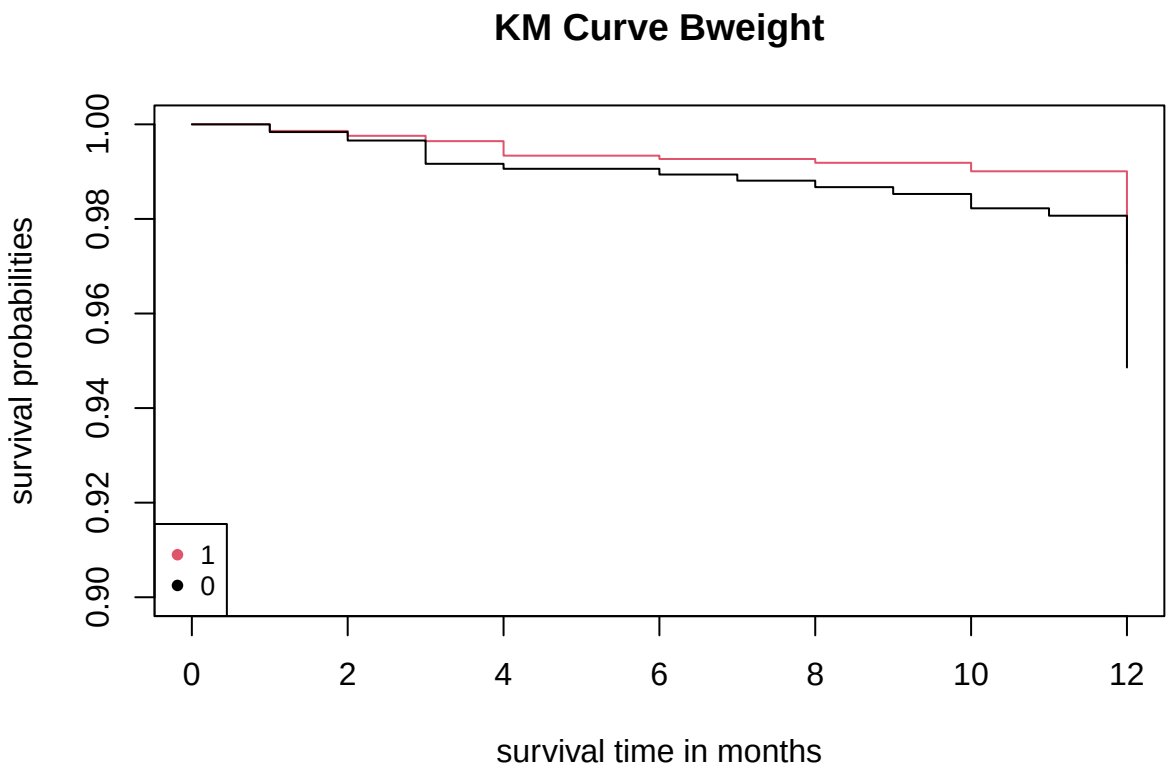

```
# region
kmfit6 <- survfit(y ~ region, data = pneumon)
plot(kmfit6, col = c(unique(pneumon$region)),xlab="survival time in months", ylab="survival probabiliti
     ylim = c(0.9,1))
title(main = "KM Curve Region")
legend("bottomleft", legend = c(unique(pneumon$region)), col=c(unique(pneumon$region)), pch = 16,
       cex = 0.8, xjust = 1)
```



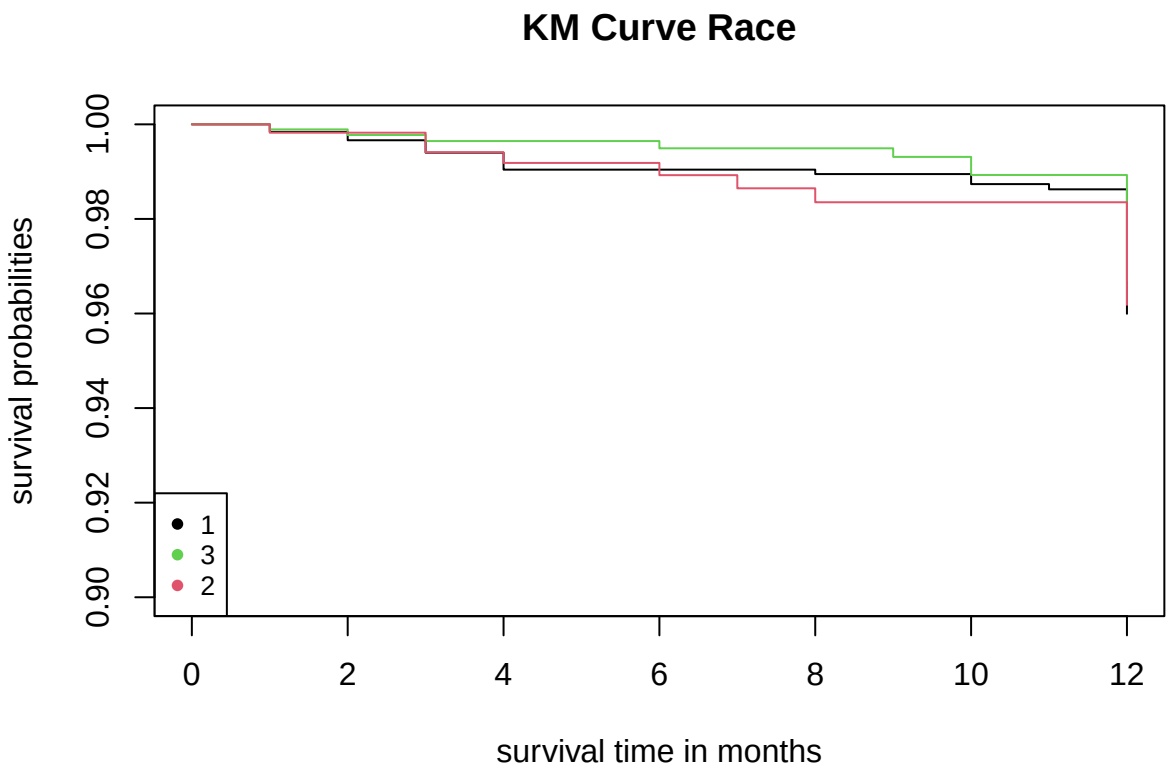

```r
# poverty
kmfit7 <- survfit(y ~ poverty, data = pneumon)
plot(kmfit7, col = c(unique(pneumon$poverty)),xlab="survival time in months", ylab="survival probabilit
     ylim = c(0.9,1))
title(main = "KM Curve Poverty")
legend("bottomleft", legend = c(unique(pneumon$poverty)), col=c(unique(pneumon$poverty)), pch = 16,
       cex = 0.8, xjust = 1)
```



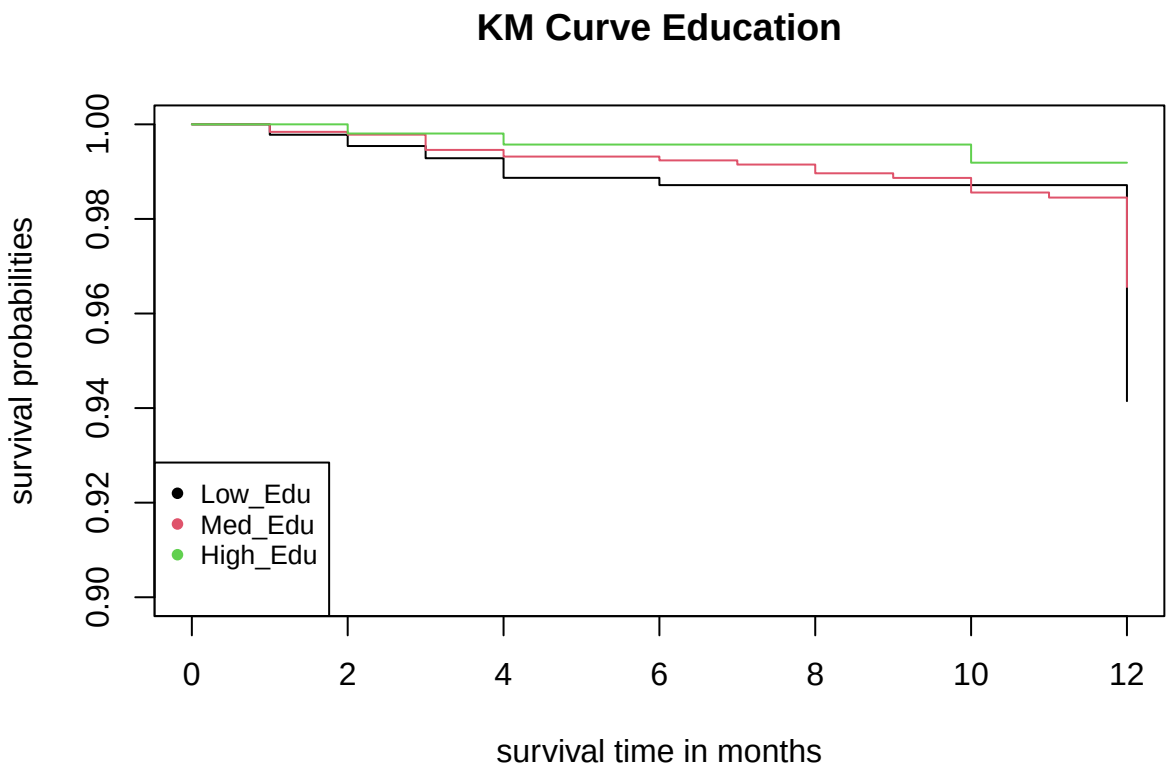

```
# bweight
kmfit8 <- survfit(y ~ bweight, data = pneumon)
plot(kmfit8, col = c(unique(pneumon$bweight)),xlab="survival time in months", ylab="survival probabilit
     ylim = c(0.9,1))
title(main = "KM Curve Bweight")
legend("bottomleft", legend = c(unique(pneumon$bweight)), col=c(unique(pneumon$bweight)), pch = 16,
       cex = 0.8, xjust = 1)
```



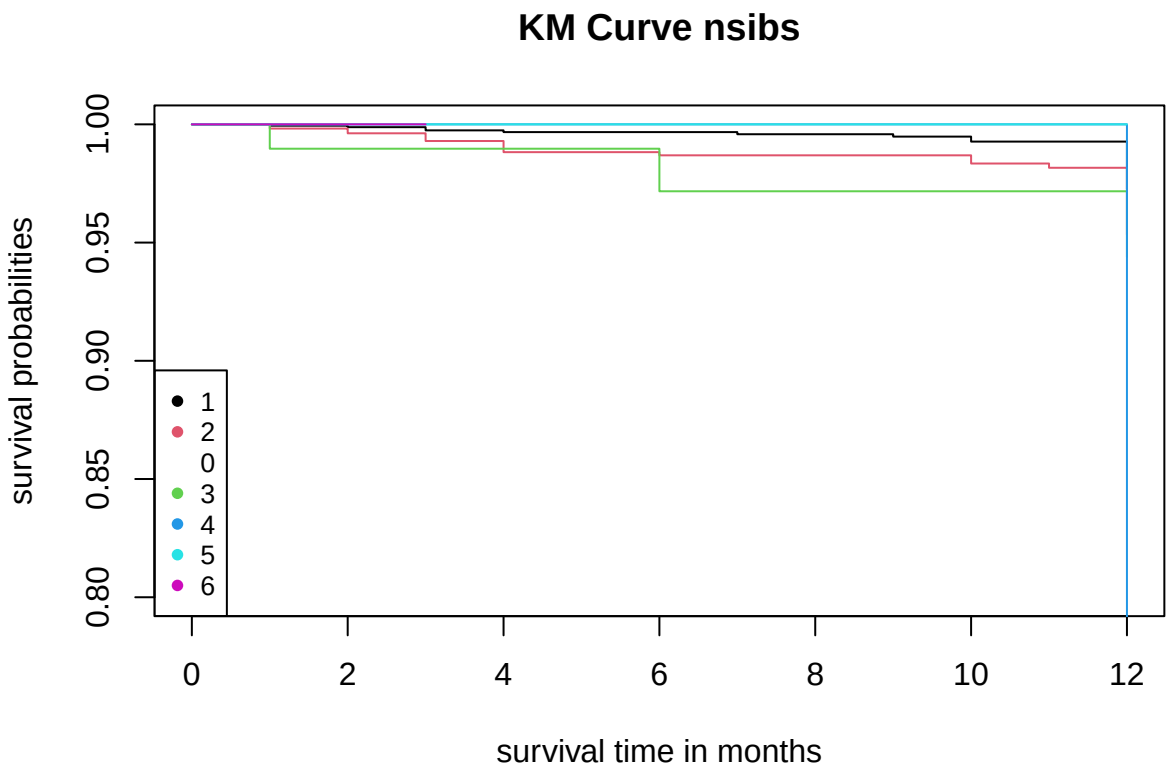

```
# race
kmfit9 <- survfit(y ~ race, data = pneumon)
plot(kmfit9, col = c(unique(pneumon$race)),xlab="survival time in months", ylab="survival probabilities"
     ylim = c(0.9,1))
title(main = "KM Curve Race")
legend("bottomleft", legend = c(unique(pneumon$race)), col=c(unique(pneumon$race)), pch = 16,
       cex = 0.8, xjust = 1)
```



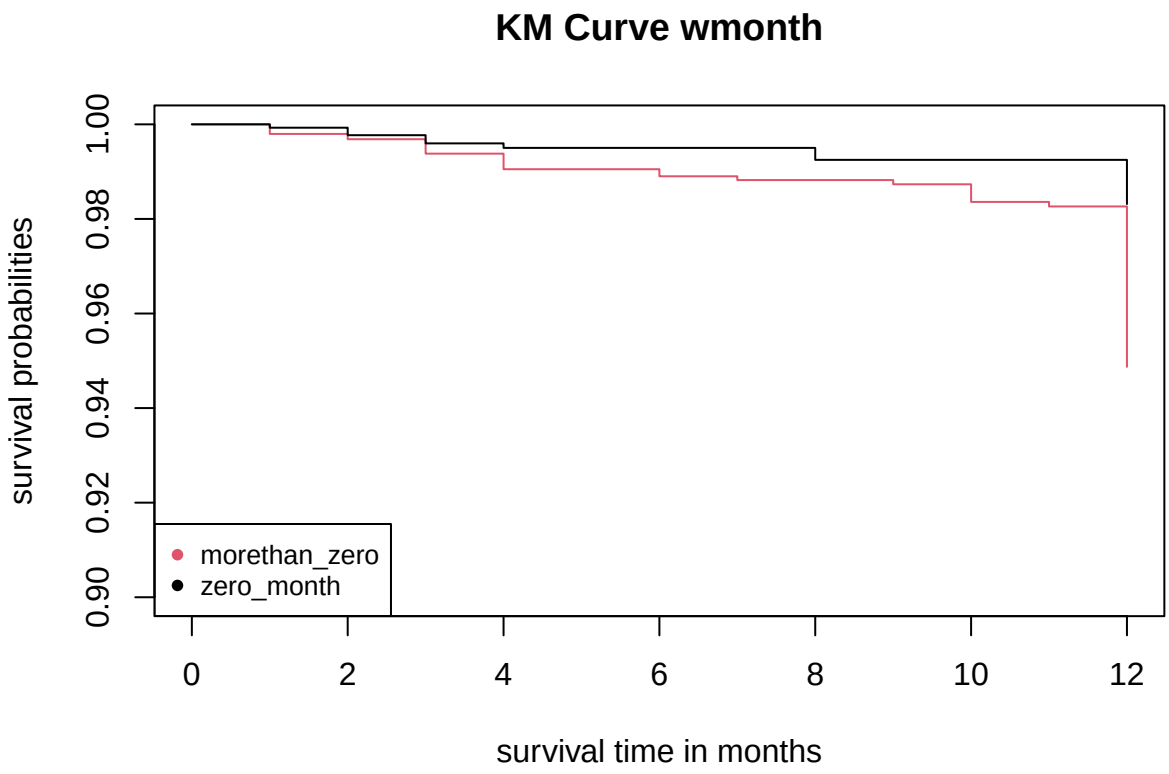

```
# education
kmfit10 <- survfit(y ~ education_cat, data = pneumon)
plot(kmfit10, col = c(unique(pneumon$education_cat)),xlab="survival time in months", ylab="survival prob
     ylim = c(0.9,1))
title(main = "KM Curve Education")
legend("bottomleft", legend = c(unique(pneumon$education_cat)), col=c(unique(pneumon$education_cat)), p
       cex = 0.8, xjust = 1)
```



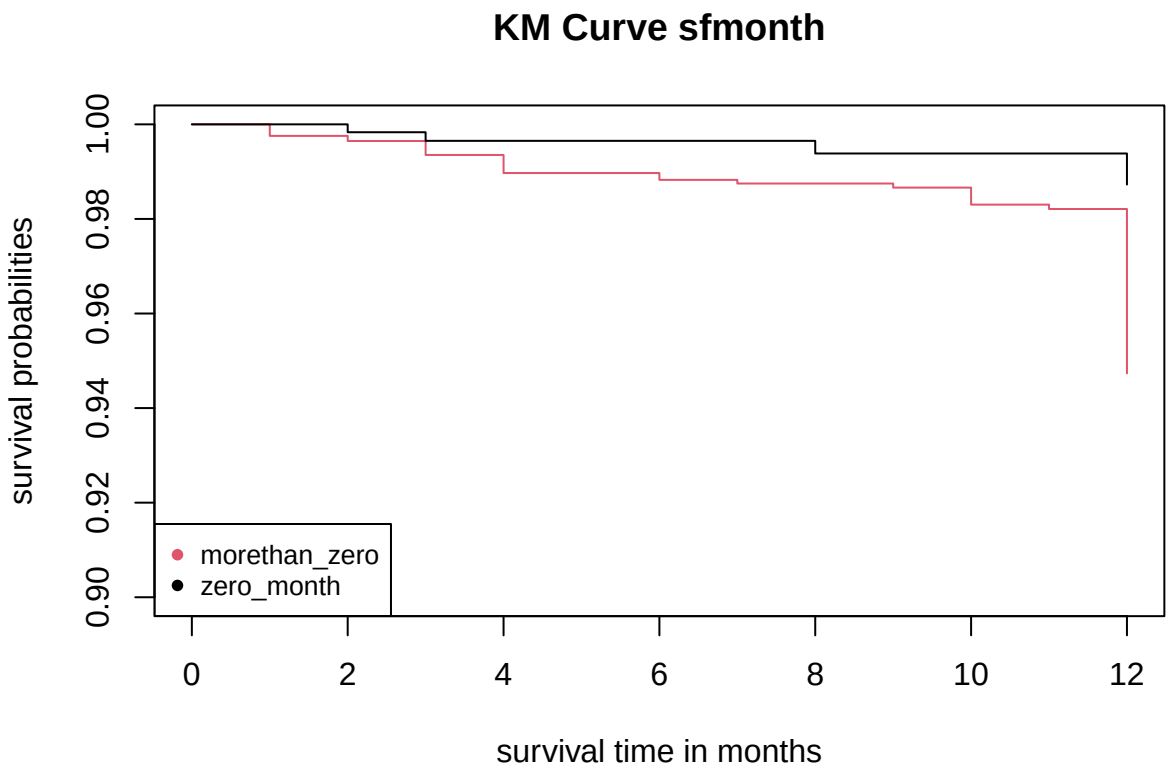

```
# nsibs
kmfit11 <- survfit(y ~ nsibs, data = pneumon)
plot(kmfit11, col = c(unique(pneumon$nsibs)),xlab="survival time in months", ylab="survival probabiliti
     ylim = c(0.8,1))
title(main = "KM Curve nsibs")
legend("bottomleft", legend = c(unique(pneumon$nsibs)), col=c(unique(pneumon$nsibs)), pch = 16,
       cex = 0.8, xjust = 1)
```




```
# wmonth
kmfit12 <- survfit(y ~ wmonth_cat, data = pneumon)
plot(kmfit12, col = c(unique(pneumon$wmonth_cat)),xlab="survival time in months", ylab="survival probabi
     ylim = c(0.9,1))
title(main = "KM Curve wmonth")
legend("bottomleft", legend = c(unique(pneumon$wmonth_cat)), col=c(unique(pneumon$wmonth_cat)), pch = 1
       cex = 0.8, xjust = 1)
```




```
# sfmonth
kmfit13 <- survfit(y ~ sfmonth_cat, data = pneumon)
plot(kmfit13, col = c(unique(pneumon$sfmonth_cat)),xlab="survival time in months", ylab="survival probal
     ylim = c(0.9,1))
title(main = "KM Curve sfmonth")
legend("bottomleft", legend = c(unique(pneumon$sfmonth_cat)), col=c(unique(pneumon$sfmonth_cat)), pch =
       cex = 0.8, xjust = 1)
```



Log Rank Test

H_0 = *The survival curves between groups are identical*

H_1 = *The survival curves between groups are not identical*

```
# chldage
LR1 <- survdiff(y ~ pneumon$chldage)
LR1$pvalue
```

```
## [1] 1.882157e-259
```

```
if (LR1$pvalue < 0.05){
  cat('Reject H0')
}else{
  cat('Fail to Reject H0')
}
```

```
## Reject H0
```

```
# mthage
LR2 <- survdiff(y ~ pneumon$mthage)
LR2$pvalue
```

```
## [1] 0.4226795
```

```r
if (LR2$pvalue < 0.05){
  cat('Reject H0')
}else{
  cat('Fail to Reject H0')
}
```

```
## Fail to Reject H0
```

```r
# urban
LR3 <- survdiff(y ~ pneumon$urban)
LR3$pvalue
```

```
## [1] 0.0624522
```

```r
if (LR3$pvalue < 0.05){
  cat('Reject H0')
}else{
  cat('Fail to Reject H0')
}
```

```
## Fail to Reject H0
```

```r
# alcohol
LR4 <- survdiff(y ~ pneumon$alcohol)
LR4$pvalue
```

```
## [1] 0.938363
```

```r
if (LR4$pvalue < 0.05){
  cat('Reject H0')
}else{
  cat('Fail to Reject H0')
}
```

```
## Fail to Reject H0
```

```r
# smoke
LR5 <- survdiff(y ~ pneumon$smoke)
LR5$pvalue
```

```
## [1] 0.0004448992
```

```r
if (LR5$pvalue < 0.05){
  cat('Reject H0')
}else{
  cat('Fail to Reject H0')
}
```

```
## Reject H0
```

```r
# region
LR6 <- survdiff(y ~ pneumon$region)
LR6$pvalue
```

```
## [1] 0.1393504
```

```r
if (LR6$pvalue < 0.05){
  cat('Reject H0')
}else{
  cat('Fail to Reject H0')
}
```

```
## Fail to Reject H0
```

```r
# poverty
LR7 <- survdiff(y ~ pneumon$poverty)
LR7$pvalue
```

```
## [1] 0.4695795
```

```r
if (LR7$pvalue < 0.05){
  cat('Reject H0')
}else{
  cat('Fail to Reject H0')
}
```

```
## Fail to Reject H0
```

```r
# bweight
LR8 <- survdiff(y ~ pneumon$bweight)
LR8$pvalue
```

```
## [1] 0.01635119
```

```r
if (LR8$pvalue < 0.05){
  cat('Reject H0')
}else{
  cat('Fail to Reject H0')
}
```

```
## Reject H0
```

```r
# race
LR9 <- survdiff(y ~ pneumon$race)
LR9$pvalue
```

```
## [1] 0.671782
```

```r
if (LR9$pvalue < 0.05){
  cat('Reject H0')
}else{
  cat('Fail to Reject H0')
}
```

```
## Fail to Reject H0
```

```r
# education
LR10 <- survdiff(y ~ pneumon$education_cat)
LR10$pvalue
```

```
## [1] 0.001892423
```

```r
if (LR10$pvalue < 0.05){
  cat('Reject H0')
}else{
  cat('Fail to Reject H0')
}
```

```
## Reject H0
```

```r
# nsibs
LR11 <- survdiff(y ~ pneumon$nsibs)
LR11$pvalue
```

```
## [1] 0.00711593
```

```r
if (LR11$pvalue < 0.05){
  cat('Reject H0')
}else{
  cat('Fail to Reject H0')
}
```

```
## Reject H0
```

```r
# wmonth
LR12 <- survdiff(y ~ pneumon$wmonth_cat)
LR12$pvalue
```

```
## [1] 0.0002569817
```

```r
if (LR12$pvalue < 0.05){
  cat('Reject H0')
}else{
  cat('Fail to Reject H0')
}
```

```
## Reject H0
```

```r
# sfmont
LR13 <- survdiff(y ~ pneumon$sfmonth_cat)
LR13$pvalue
```

```
## [1] 1.662848e-05
```

```r
if (LR13$pvalue < 0.05){
  cat('Reject H0')
}else{
  cat('Fail to Reject H0')
}
```

```
## Reject H0
```

Checking Proportional Hazard Assumptions

Goodness of Fit

H_0 = *Fulfill PH assumption*

H_1 = *PH asusmption not fulfilled*

```r
model1 <- coxph(y~. -agepn -hospital, data = pneumon)
summary(model1)
```

```
## Call:
## coxph(formula = y ~ . - agepn - hospital, data = pneumon)
##
##   n= 3469, number of events= 73
##    (1 observation deleted due to missingness)
##
##                               coef exp(coef) se(coef)       z Pr(>|z|)
## chldage                   -0.58057   0.55958  0.03572 -16.253  < 2e-16 ***
## mthage                    -0.24543   0.78237  0.07074  -3.470 0.000521 ***
## urban1                    -0.29416   0.74516  0.30993  -0.949 0.342567
## alcohol1                   0.29842   1.34772  0.27294   1.093 0.274241
## smoke1                     0.22358   1.25054  0.29803   0.750 0.453148
## region2                    0.49810   1.64560  0.41733   1.194 0.232660
## region3                    0.12657   1.13493  0.42703   0.296 0.766923
## region4                   -0.48939   0.61300  0.47904  -1.022 0.306975
## poverty1                  -0.08711   0.91658  0.44476  -0.196 0.844726
## bweight1                   0.08945   1.09357  0.29003   0.308 0.757772
## race2                     -0.08382   0.91959  0.31154  -0.269 0.787880
## race3                      0.43249   1.54109  0.38755   1.116 0.264440
## nsibs                      0.31667   1.37255  0.16188   1.956 0.050440 .
## education_catMed_Edu       0.27346   1.31451  0.29781   0.918 0.358486
## education_catHigh_Edu      0.49865   1.64650  0.70105   0.711 0.476902
## wmonth_catmorethan_zero    1.02306   2.78168  0.61579   1.661 0.096639 .
## sfmonth_catmorethan_zero -1.33576   0.26296  0.67202  -1.988 0.046847 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
##                           exp(coef) exp(-coef) lower .95 upper .95
## chldage                      0.5596     1.7871   0.52174    0.6002
## mthage                       0.7824     1.2782   0.68108    0.8987
## urban1                       0.7452     1.3420   0.40592    1.3679
## alcohol1                     1.3477     0.7420   0.78936    2.3011
## smoke1                       1.2505     0.7997   0.69729    2.2428
## region2                      1.6456     0.6077   0.72625    3.7287
## region3                      1.1349     0.8811   0.49145    2.6210
## region4                      0.6130     1.6313   0.23972    1.5676
## poverty1                     0.9166     1.0910   0.38335    2.1915
## bweight1                     1.0936     0.9144   0.61940    1.9307
## race2                        0.9196     1.0874   0.49936    1.6935
## race3                        1.5411     0.6489   0.72101    3.2939
## nsibs                        1.3725     0.7286   0.99939    1.8850
## education_catMed_Edu         1.3145     0.7607   0.73328    2.3565
## education_catHigh_Edu        1.6465     0.6073   0.41671    6.5057
## wmonth_catmorethan_zero      2.7817     0.3595   0.83204    9.2997
## sfmonth_catmorethan_zero     0.2630     3.8029   0.07045    0.9815
##
## Concordance= 0.967  (se = 0.01 )
## Likelihood ratio test= 413.5  on 17 df,   p=<2e-16
## Wald test            = 311.4  on 17 df,   p=<2e-16
## Score (logrank) test = 1204  on 17 df,   p=<2e-16
```

```
check_ph <- cox.zph(model1, transform = rank)
check_ph$table
```

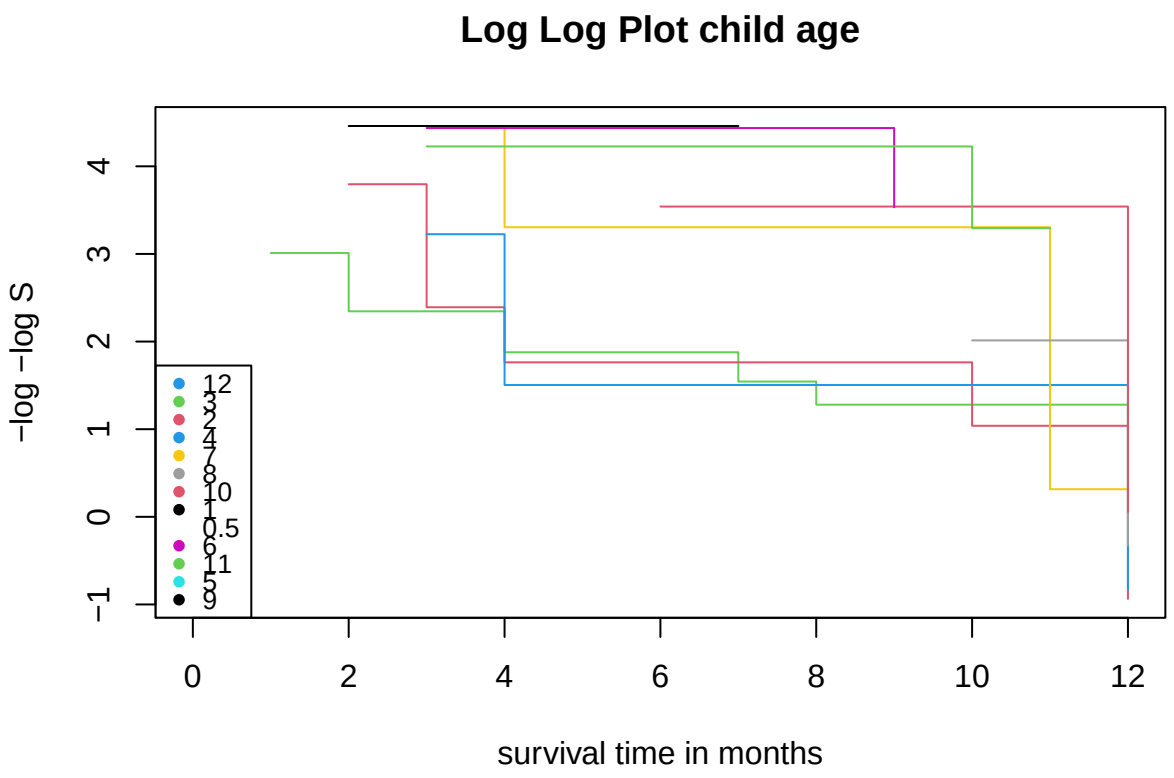
```
##                       chisq df         p
## chldage       2.352362e-01  1 0.6276682
## mthage        8.137652e-01  1 0.3670094
## urban         1.175090e+00  1 0.2783579
## alcohol       3.251151e-02  1 0.8569096
## smoke         9.110215e-03  1 0.9239594
## region        4.371091e-01  3 0.9324743
## poverty       1.173098e+00  1 0.2787657
## bweight       1.236974e-03  1 0.9719437
## race          9.325947e-01  2 0.6273207
## nsibs         1.159256e+00  1 0.2816198
## education_cat 6.564873e-02  2 0.9677085
## wmonth_cat    1.969511e-06  1 0.9988803
## sfmonth_cat   3.651496e-02  1 0.8484560
## GLOBAL        5.785092e+00 17 0.9945172
```

Log Log Plot

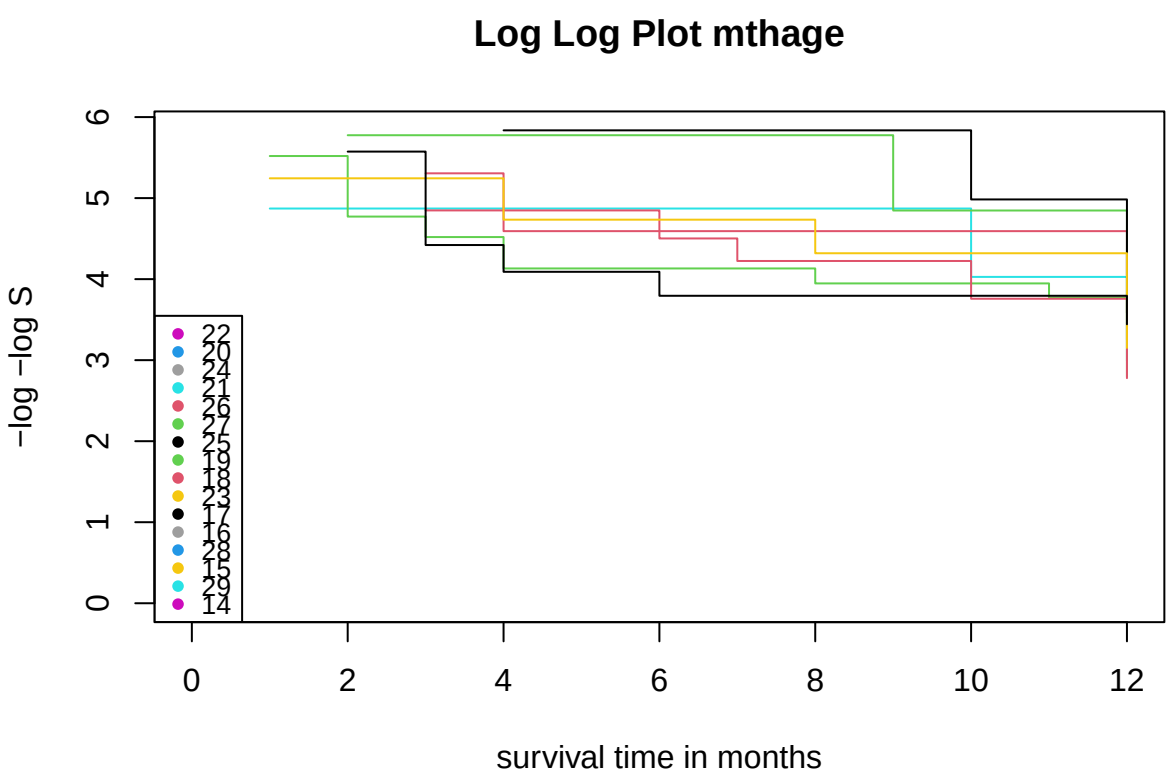
```
minusloglog <- function(p){
  return(-log(-log(p)))
}

# child age
win.graph()
plot(kmfit1, fun = minusloglog, col = c(unique(pneumon$chldage)),xlab="survival time in months", ylab="
title(main = "Log Log Plot child age")
```

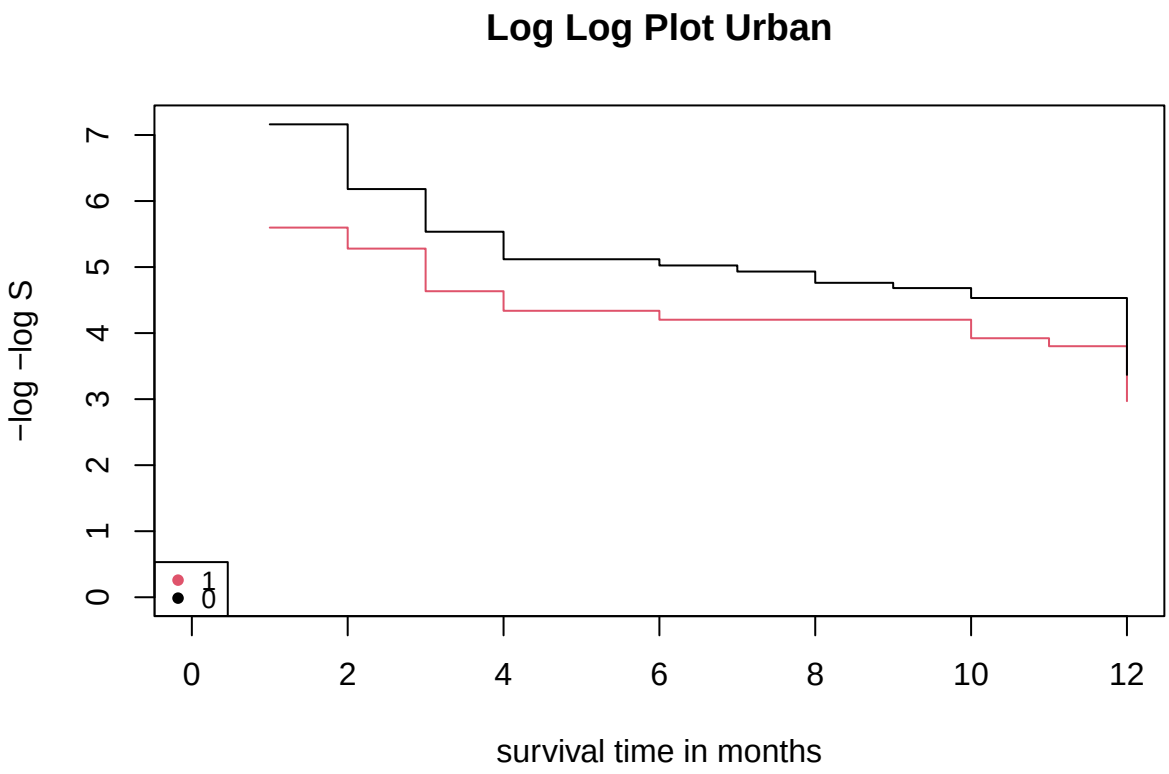
```
legend("bottomleft", legend = c(unique(pneumon$chldage)), col=c(unique(pneumon$chldage)), pch = 16,
       cex = 0.8, xjust = 1)
```



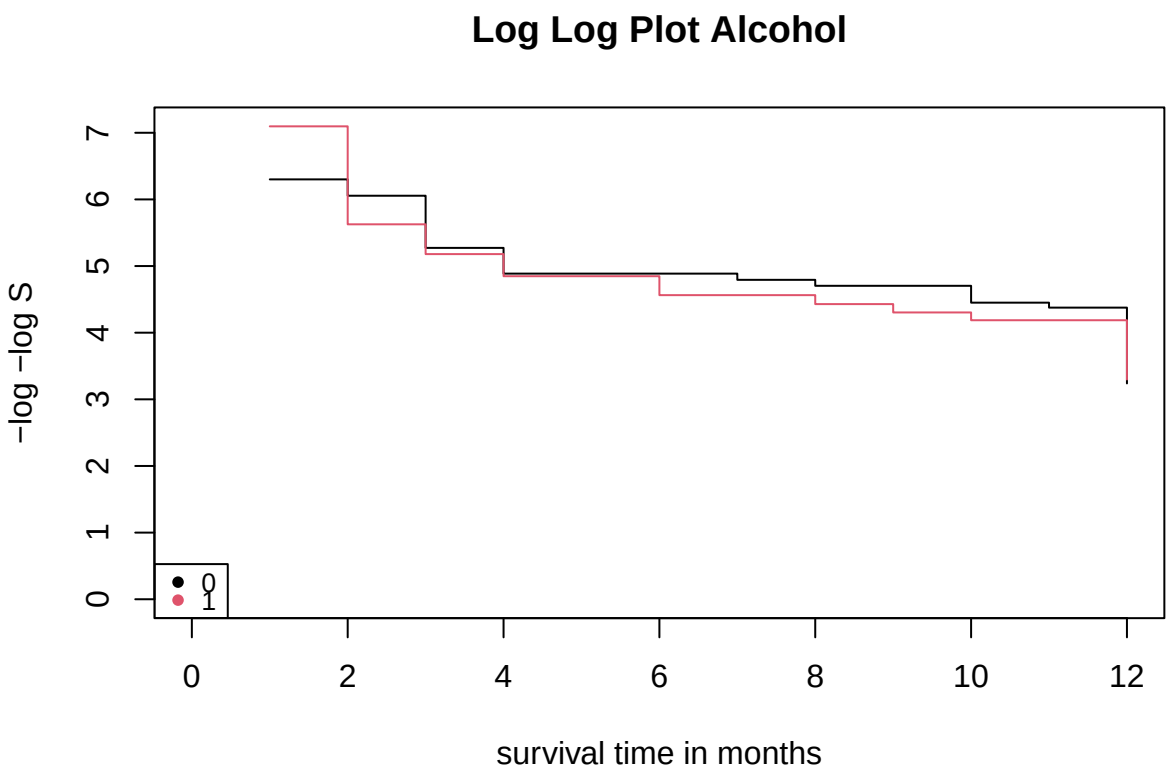

```
# mthage
plot(kmfit2, fun = minusloglog, col = c(unique(pneumon$mthage)),xlab="survival time in months", ylab="-l
title(main = "Log Log Plot mthage")
legend("bottomleft", legend = c(unique(pneumon$mthage)), col=c(unique(pneumon$mthage)), pch = 16,
       cex = 0.8, xjust = 1)
```



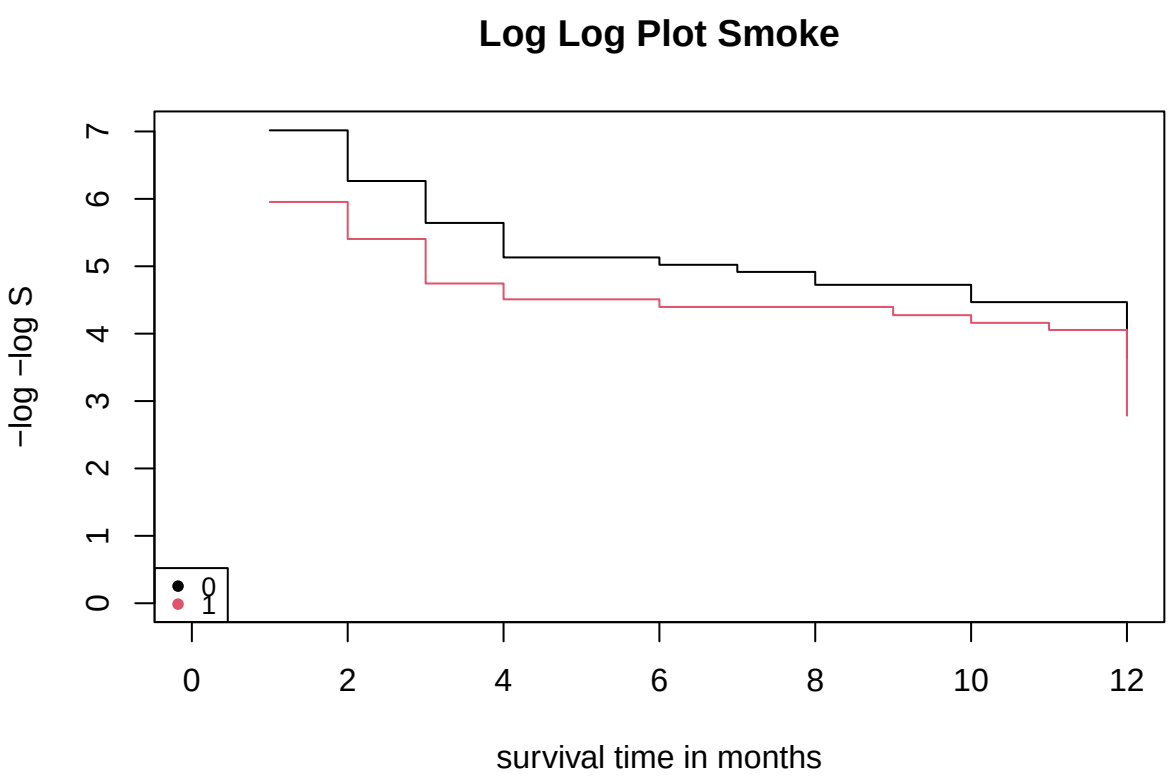

```
# urban
plot(kmfit3, fun = minusloglog, col = c(unique(pneumon$urban)),xlab="survival time in months", ylab="-l
title(main = "Log Log Plot Urban")
legend("bottomleft", legend = c(unique(pneumon$urban)), col=c(unique(pneumon$urban)), pch = 16,
       cex = 0.8, xjust = 1)
```



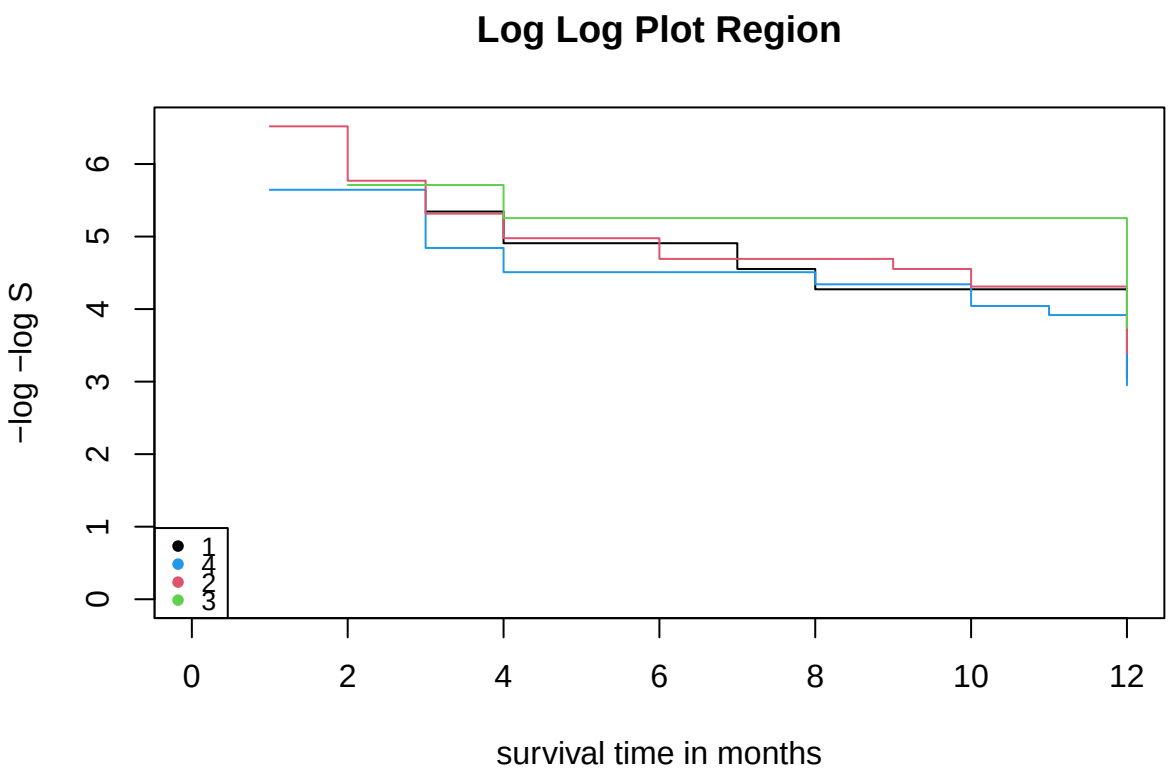

```
# alchohol
plot(kmfit4, fun = minusloglog, col = c(unique(pneumon$alcohol)),xlab="survival time in months", ylab="
title(main = "Log Log Plot Alcohol")
legend("bottomleft", legend = c(unique(pneumon$alcohol)), col=c(unique(pneumon$alcohol)), pch = 16,
       cex = 0.8, xjust = 1)
```



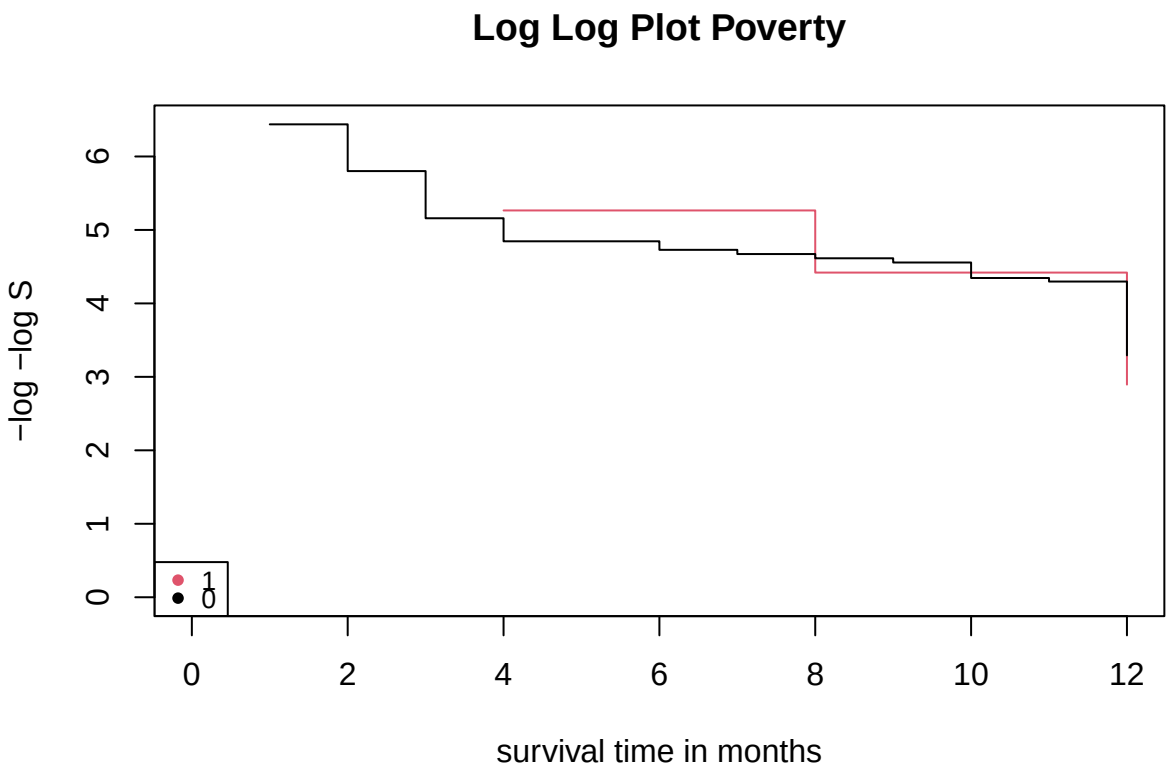

```
# smoke
plot(kmfit5, fun = minusloglog, col = c(unique(pneumon$smoke)),xlab="survival time in months", ylab="-l
title(main = "Log Log Plot Smoke")
legend("bottomleft", legend = c(unique(pneumon$smoke)), col=c(unique(pneumon$smoke)), pch = 16,
       cex = 0.8, xjust = 1)
```



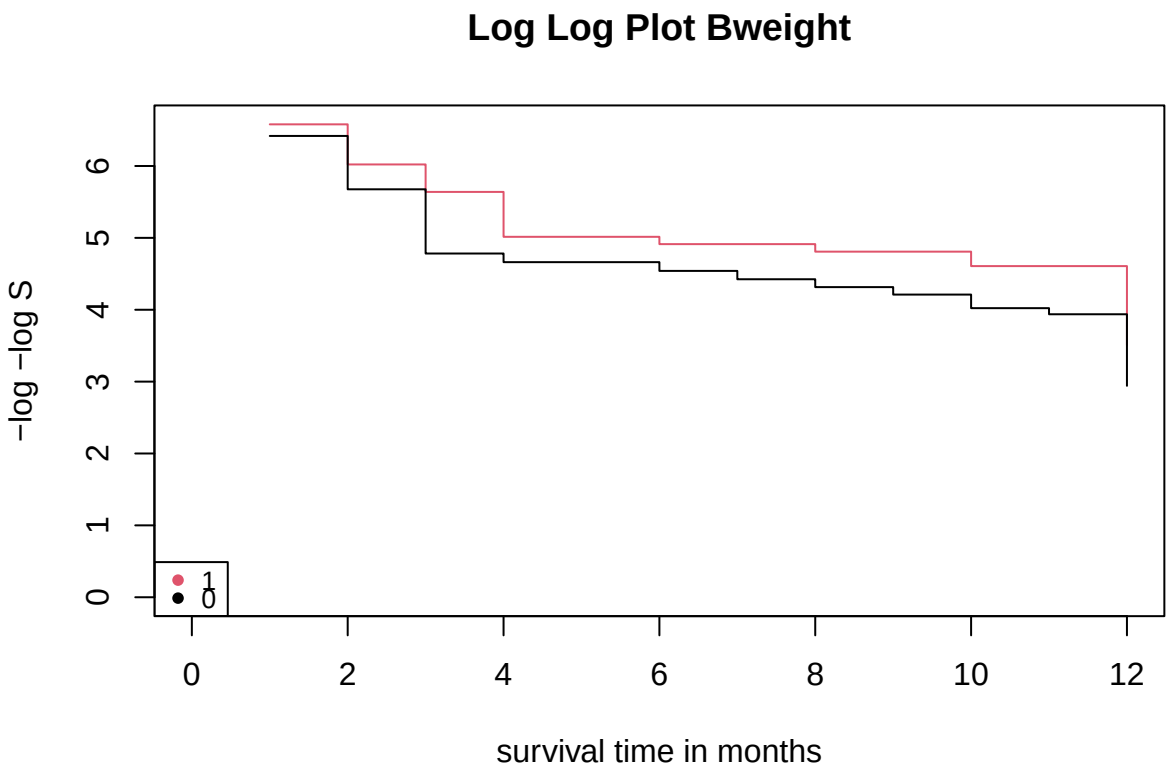

```
# region
plot(kmfit6, fun = minusloglog, col = c(unique(pneumon$region)),xlab="survival time in months", ylab="-l
title(main = "Log Log Plot Region")
legend("bottomleft", legend = c(unique(pneumon$region)), col=c(unique(pneumon$region)), pch = 16,
       cex = 0.8, xjust = 1)
```



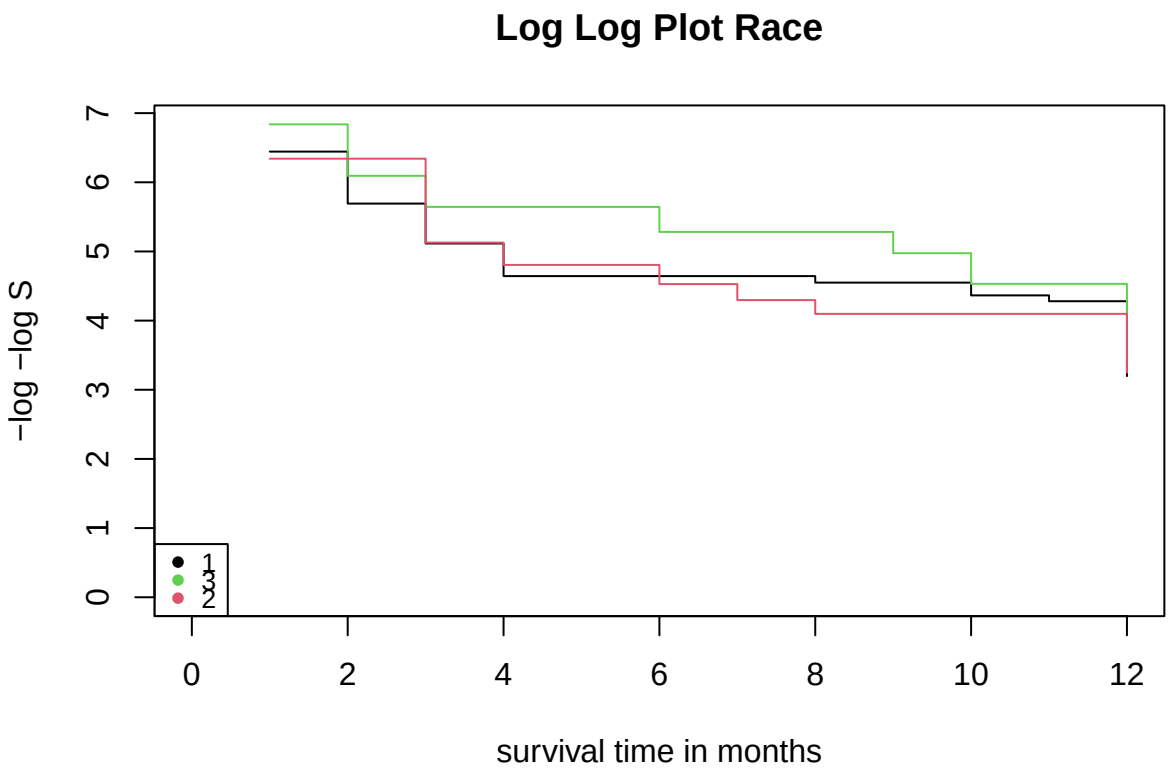

```r
# poverty
plot(kmfit7, fun = minusloglog, col = c(unique(pneumon$poverty)),xlab="survival time in months", ylab="
title(main = "Log Log Plot Poverty")
legend("bottomleft", legend = c(unique(pneumon$poverty)), col=c(unique(pneumon$poverty)), pch = 16,
       cex = 0.8, xjust = 1)
```



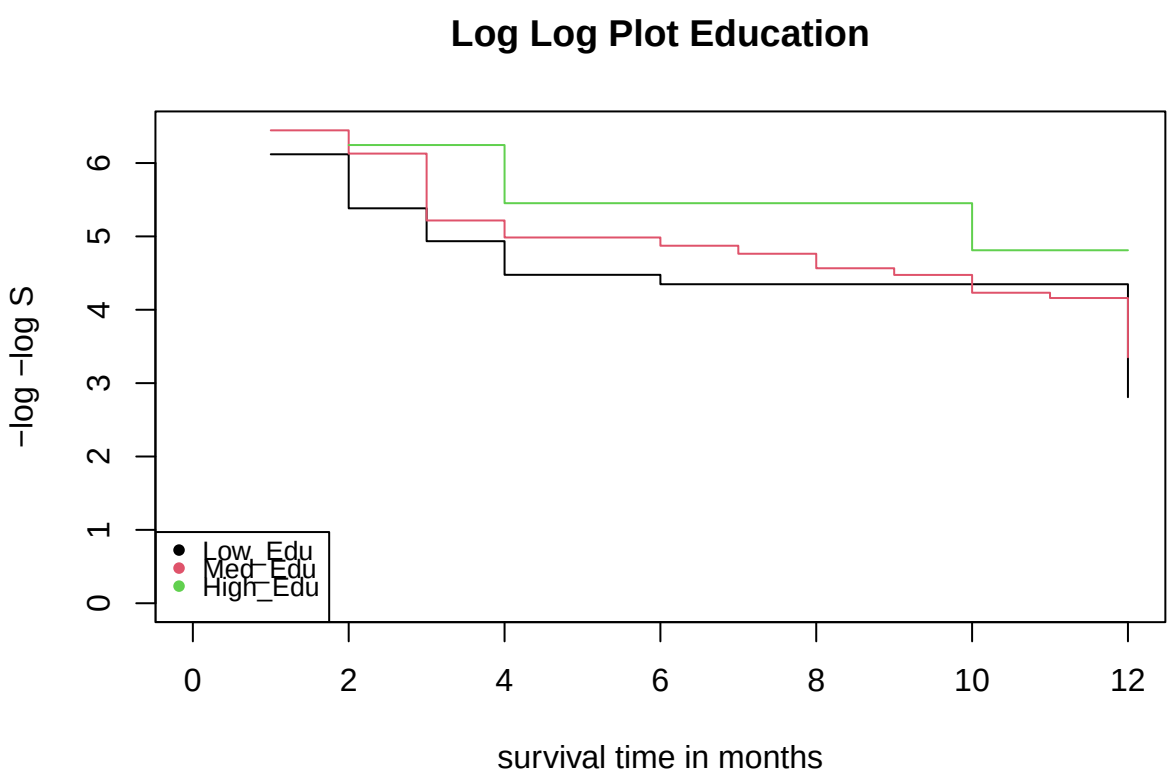

```r
# bweight
plot(kmfit8, fun = minusloglog, col = c(unique(pneumon$bweight)),xlab="survival time in months", ylab="
title(main = "Log Log Plot Bweight")
legend("bottomleft", legend = c(unique(pneumon$bweight)), col=c(unique(pneumon$bweight)), pch = 16,
       cex = 0.8, xjust = 1)
```



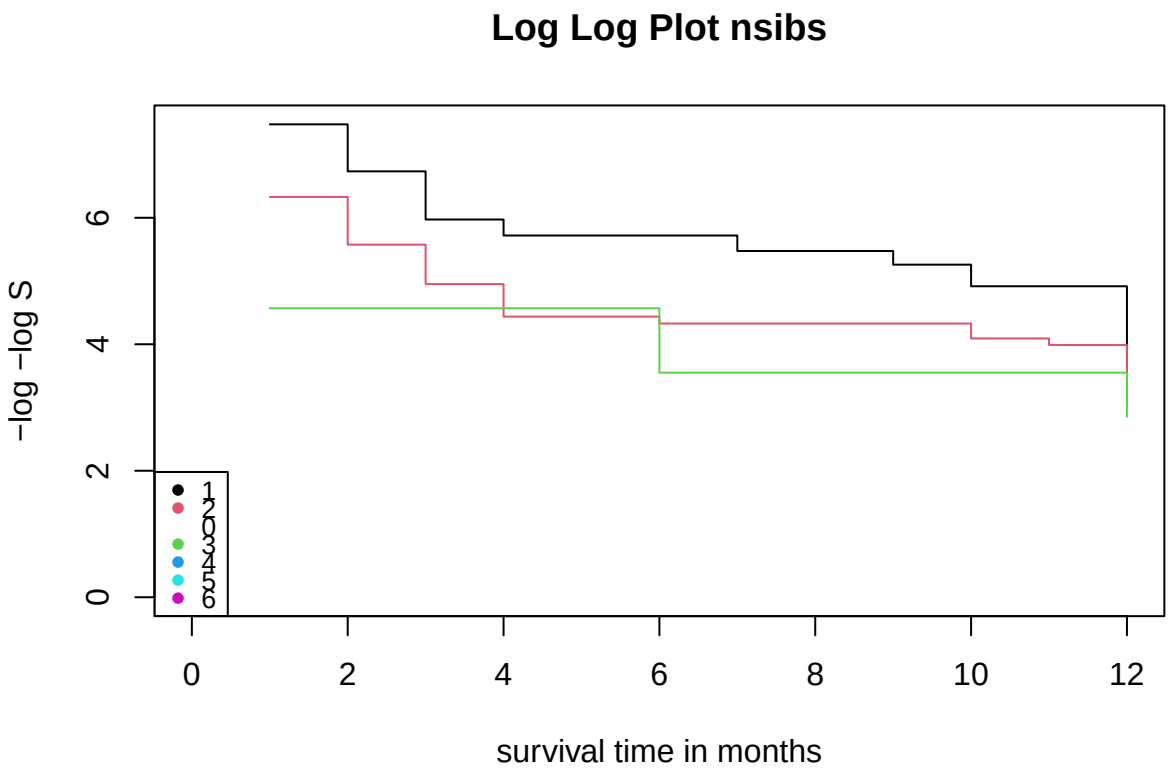

```r
# race
plot(kmfit9, fun = minusloglog, col = c(unique(pneumon$race)),xlab="survival time in months", ylab="-log
title(main = "Log Log Plot Race")
legend("bottomleft", legend = c(unique(pneumon$race)), col=c(unique(pneumon$race)), pch = 16,
       cex = 0.8, xjust = 1)
```



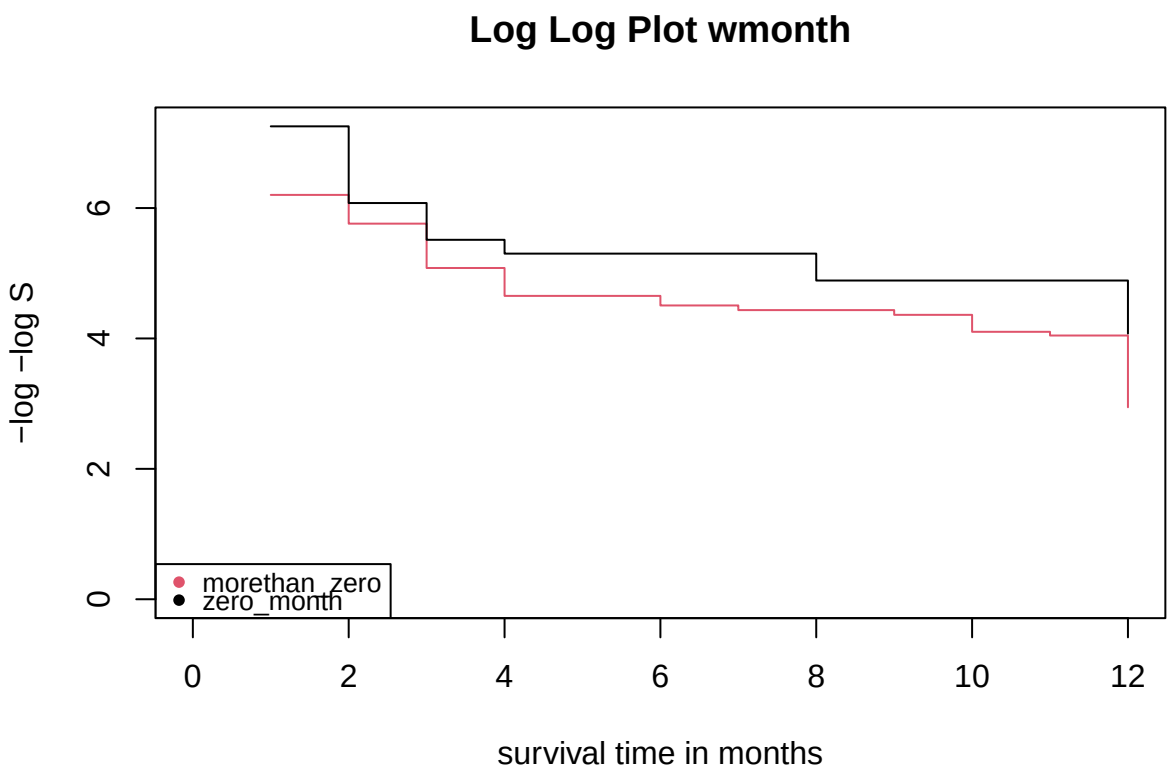

```
# education
plot(kmfit10, fun = minusloglog, col = c(unique(pneumon$education_cat)),xlab="survival time in months",
title(main = "Log Log Plot Education")
legend("bottomleft", legend = c(unique(pneumon$education_cat)), col=c(unique(pneumon$education_cat)), p
       cex = 0.8, xjust = 1)
```



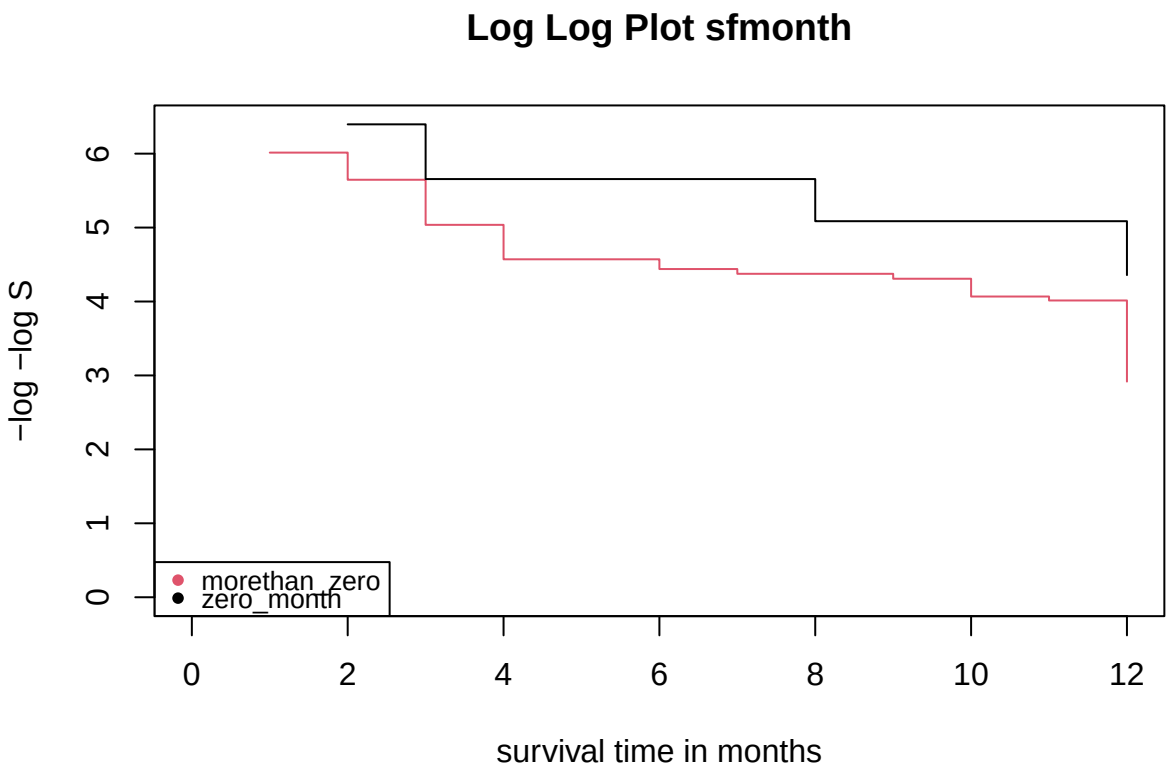

```
# nsibs
plot(kmfit11, fun = minusloglog, col = c(unique(pneumon$nsibs)),xlab="survival time in months", ylab="-l
title(main = "Log Log Plot nsibs")
legend("bottomleft", legend = c(unique(pneumon$nsibs)), col=c(unique(pneumon$nsibs)), pch = 16,
       cex = 0.8, xjust = 1)
```




```
# wmonth
plot(kmfit12, fun = minusloglog, col = c(unique(pneumon$wmonth_cat)),xlab="survival time in months", yla
title(main = "Log Log Plot wmonth")
legend("bottomleft", legend = c(unique(pneumon$wmonth_cat)), col=c(unique(pneumon$wmonth_cat)), pch = 1
       cex = 0.8, xjust = 1)
```




```
# sfmonth
plot(kmfit13, fun = minusloglog, col = c(unique(pneumon$sfmonth_cat)),xlab="survival time in months", yl
title(main = "Log Log Plot sfmonth")
legend("bottomleft", legend = c(unique(pneumon$sfmonth_cat)), col=c(unique(pneumon$sfmonth_cat)), pch =
       cex = 0.8, xjust = 1)
```


Log Log Plot sfmonth
−log −log S
0
1
2
3
4
5
6
morethan_zero
zero_month
0
2
4
6
8
10
12
survival time in months