problem-set-2-term-3.R

USER

2024-02-20

library(tidyverse)

## Warning: package 'ggplot2' was built under R version 4.3.2

## Warning: package 'dplyr' was built under R version 4.3.2

## Warning: package 'lubridate' was built under R version 4.3.2

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.3 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.4.4 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(survival)

## Warning: package 'survival' was built under R version 4.3.2

library(ggsurvfit)

## Warning: package 'ggsurvfit' was built under R version 4.3.2

library(finalfit)

## Warning: package 'finalfit' was built under R version 4.3.2

library(broom)  
library(gtsummary)

## Warning: package 'gtsummary' was built under R version 4.3.2

## #BlackLivesMatter

library(ggeasy)

## Warning: package 'ggeasy' was built under R version 4.3.2

library(survminer)

## Warning: package 'survminer' was built under R version 4.3.2

## Loading required package: ggpubr

## Warning: package 'ggpubr' was built under R version 4.3.2

##   
## Attaching package: 'survminer'  
##   
## The following object is masked from 'package:survival':  
##   
## myeloma

# question a --------------------------------------------------------------  
  
pbcData <- read\_csv("pbctrial.csv")

## Rows: 312 Columns: 18  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (2): sex, agecat  
## dbl (16): case, drug, bil, histo, death, survyr, \_st, \_d, \_t, \_t0, ageyr, ag...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

ff\_glimpse(pbcData)

## $Continuous  
## label var\_type n missing\_n missing\_percent mean sd min  
## case case <dbl> 312 0 0.0 156.5 90.2 1.0  
## drug drug <dbl> 312 0 0.0 0.5 0.5 0.0  
## bil bil <dbl> 312 0 0.0 3.3 4.5 0.3  
## histo histo <dbl> 312 0 0.0 3.0 0.9 1.0  
## death death <dbl> 312 0 0.0 0.4 0.5 0.0  
## survyr survyr <dbl> 312 0 0.0 5.5 3.1 0.1  
## X\_st X\_st <dbl> 312 0 0.0 1.0 0.0 1.0  
## X\_d X\_d <dbl> 312 0 0.0 0.4 0.5 0.0  
## X\_t X\_t <dbl> 312 0 0.0 5.5 3.1 0.1  
## X\_t0 X\_t0 <dbl> 312 0 0.0 0.0 0.0 0.0  
## ageyr ageyr <dbl> 312 0 0.0 50.1 10.6 26.3  
## agegr\_2 agegr\_2 <dbl> 312 0 0.0 0.3 0.5 0.0  
## agegr\_3 agegr\_3 <dbl> 312 0 0.0 0.3 0.5 0.0  
## hstage2 hstage2 <dbl> 312 0 0.0 0.2 0.4 0.0  
## hstage3 hstage3 <dbl> 312 0 0.0 0.4 0.5 0.0  
## hstage4 hstage4 <dbl> 312 0 0.0 0.3 0.5 0.0  
## quartile\_25 median quartile\_75 max  
## case 78.8 156.5 234.2 312.0  
## drug 0.0 1.0 1.0 1.0  
## bil 0.8 1.4 3.4 28.0  
## histo 2.0 3.0 4.0 4.0  
## death 0.0 0.0 1.0 1.0  
## survyr 3.3 5.0 7.4 12.5  
## X\_st 1.0 1.0 1.0 1.0  
## X\_d 0.0 0.0 1.0 1.0  
## X\_t 3.3 5.0 7.4 12.5  
## X\_t0 0.0 0.0 0.0 0.0  
## ageyr 42.3 49.8 56.8 78.5  
## agegr\_2 0.0 0.0 1.0 1.0  
## agegr\_3 0.0 0.0 1.0 1.0  
## hstage2 0.0 0.0 0.0 1.0  
## hstage3 0.0 0.0 1.0 1.0  
## hstage4 0.0 0.0 1.0 1.0  
##   
## $Categorical  
## label var\_type n missing\_n missing\_percent levels\_n levels  
## sex sex <chr> 312 0 0.0 2 -  
## agecat agecat <chr> 312 0 0.0 3 -  
## levels\_count levels\_percent  
## sex - -  
## agecat - -

summary(pbcData)

## case drug sex bil   
## Min. : 1.00 Min. :0.0000 Length:312 Min. : 0.300   
## 1st Qu.: 78.75 1st Qu.:0.0000 Class :character 1st Qu.: 0.800   
## Median :156.50 Median :1.0000 Mode :character Median : 1.350   
## Mean :156.50 Mean :0.5064 Mean : 3.256   
## 3rd Qu.:234.25 3rd Qu.:1.0000 3rd Qu.: 3.425   
## Max. :312.00 Max. :1.0000 Max. :28.000   
## histo death survyr \_st   
## Min. :1.000 Min. :0.0000 Min. : 0.1123 Min. :1   
## 1st Qu.:2.000 1st Qu.:0.0000 1st Qu.: 3.2630 1st Qu.:1   
## Median :3.000 Median :0.0000 Median : 5.0397 Median :1   
## Mean :3.032 Mean :0.4006 Mean : 5.4969 Mean :1   
## 3rd Qu.:4.000 3rd Qu.:1.0000 3rd Qu.: 7.3897 3rd Qu.:1   
## Max. :4.000 Max. :1.0000 Max. :12.4822 Max. :1   
## \_d \_t \_t0 ageyr   
## Min. :0.0000 Min. : 0.1123 Min. :0 Min. :26.30   
## 1st Qu.:0.0000 1st Qu.: 3.2630 1st Qu.:0 1st Qu.:42.27   
## Median :0.0000 Median : 5.0397 Median :0 Median :49.83   
## Mean :0.4006 Mean : 5.4969 Mean :0 Mean :50.05   
## 3rd Qu.:1.0000 3rd Qu.: 7.3897 3rd Qu.:0 3rd Qu.:56.75   
## Max. :1.0000 Max. :12.4822 Max. :0 Max. :78.49   
## agecat agegr\_2 agegr\_3 hstage2   
## Length:312 Min. :0.0000 Min. :0.0000 Min. :0.0000   
## Class :character 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000   
## Mode :character Median :0.0000 Median :0.0000 Median :0.0000   
## Mean :0.3237 Mean :0.3365 Mean :0.2147   
## 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:0.0000   
## Max. :1.0000 Max. :1.0000 Max. :1.0000   
## hstage3 hstage4   
## Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:0.0000   
## Median :0.0000 Median :0.0000   
## Mean :0.3846 Mean :0.3494   
## 3rd Qu.:1.0000 3rd Qu.:1.0000   
## Max. :1.0000 Max. :1.0000

str(pbcData)

## spc\_tbl\_ [312 × 18] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ case : num [1:312] 1 2 3 4 5 6 7 8 9 10 ...  
## $ drug : num [1:312] 1 1 1 1 0 0 0 0 1 0 ...  
## $ sex : chr [1:312] "Female" "Female" "Male" "Female" ...  
## $ bil : num [1:312] 14.5 1.1 1.4 1.8 3.4 ...  
## $ histo : num [1:312] 4 3 4 4 3 3 3 3 2 4 ...  
## $ death : num [1:312] 1 0 1 1 0 1 0 1 1 1 ...  
## $ survyr : num [1:312] 1.1 12.33 2.77 5.27 4.12 ...  
## $ \_st : num [1:312] 1 1 1 1 1 1 1 1 1 1 ...  
## $ \_d : num [1:312] 1 0 1 1 0 1 0 1 1 1 ...  
## $ \_t : num [1:312] 1.1 12.33 2.77 5.27 4.12 ...  
## $ \_t0 : num [1:312] 0 0 0 0 0 0 0 0 0 0 ...  
## $ ageyr : num [1:312] 58.8 56.5 70.1 54.8 38.1 ...  
## $ agecat : chr [1:312] ">= 55 yrs" ">= 55 yrs" ">= 55 yrs" "45 - 55 yrs" ...  
## $ agegr\_2: num [1:312] 0 0 0 1 0 0 0 1 0 0 ...  
## $ agegr\_3: num [1:312] 1 1 1 0 0 1 1 0 0 1 ...  
## $ hstage2: num [1:312] 0 0 0 0 0 0 0 0 1 0 ...  
## $ hstage3: num [1:312] 0 1 0 0 1 1 1 1 0 0 ...  
## $ hstage4: num [1:312] 1 0 1 1 0 0 0 0 0 1 ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. case = col\_double(),  
## .. drug = col\_double(),  
## .. sex = col\_character(),  
## .. bil = col\_double(),  
## .. histo = col\_double(),  
## .. death = col\_double(),  
## .. survyr = col\_double(),  
## .. `\_st` = col\_double(),  
## .. `\_d` = col\_double(),  
## .. `\_t` = col\_double(),  
## .. `\_t0` = col\_double(),  
## .. ageyr = col\_double(),  
## .. agecat = col\_character(),  
## .. agegr\_2 = col\_double(),  
## .. agegr\_3 = col\_double(),  
## .. hstage2 = col\_double(),  
## .. hstage3 = col\_double(),  
## .. hstage4 = col\_double()  
## .. )  
## - attr(\*, "problems")=<externalptr>

## counts  
 pbcData |>   
 select(sex,drug,death,histo,agecat) |>   
 map(~ table(.x)) |>   
 map(print)

## .x  
## Female Male   
## 276 36   
## .x  
## 0 1   
## 154 158   
## .x  
## 0 1   
## 187 125   
## .x  
## 1 2 3 4   
## 16 67 120 109   
## .x  
## < 45 yrs >= 55 yrs 45 - 55 yrs   
## 106 105 101

## $sex  
## .x  
## Female Male   
## 276 36   
##   
## $drug  
## .x  
## 0 1   
## 154 158   
##   
## $death  
## .x  
## 0 1   
## 187 125   
##   
## $histo  
## .x  
## 1 2 3 4   
## 16 67 120 109   
##   
## $agecat  
## .x  
## < 45 yrs >= 55 yrs 45 - 55 yrs   
## 106 105 101

## proportions  
pbcData |>   
 select(sex,drug,death,histo,agecat) |>   
 map(~prop.table(table(.x))) |>   
 map(print)

## .x  
## Female Male   
## 0.8846154 0.1153846   
## .x  
## 0 1   
## 0.4935897 0.5064103   
## .x  
## 0 1   
## 0.599359 0.400641   
## .x  
## 1 2 3 4   
## 0.05128205 0.21474359 0.38461538 0.34935897   
## .x  
## < 45 yrs >= 55 yrs 45 - 55 yrs   
## 0.3397436 0.3365385 0.3237179

## $sex  
## .x  
## Female Male   
## 0.8846154 0.1153846   
##   
## $drug  
## .x  
## 0 1   
## 0.4935897 0.5064103   
##   
## $death  
## .x  
## 0 1   
## 0.599359 0.400641   
##   
## $histo  
## .x  
## 1 2 3 4   
## 0.05128205 0.21474359 0.38461538 0.34935897   
##   
## $agecat  
## .x  
## < 45 yrs >= 55 yrs 45 - 55 yrs   
## 0.3397436 0.3365385 0.3237179

## continuos variables  
pbcData |>   
 select(bil,survyr,ageyr) |>   
 summary()

## bil survyr ageyr   
## Min. : 0.300 Min. : 0.1123 Min. :26.30   
## 1st Qu.: 0.800 1st Qu.: 3.2630 1st Qu.:42.27   
## Median : 1.350 Median : 5.0397 Median :49.83   
## Mean : 3.256 Mean : 5.4969 Mean :50.05   
## 3rd Qu.: 3.425 3rd Qu.: 7.3897 3rd Qu.:56.75   
## Max. :28.000 Max. :12.4822 Max. :78.49

## check for missing values in each variable  
missing\_na <- pbcData |>   
 summarise(across(everything(),~ sum(is.na(.x)))) |>   
 pivot\_longer(everything(),names\_to = "'Variable",values\_to = " Missing")  
print(missing\_na)

## # A tibble: 18 × 2  
## `'Variable` ` Missing`  
## <chr> <int>  
## 1 case 0  
## 2 drug 0  
## 3 sex 0  
## 4 bil 0  
## 5 histo 0  
## 6 death 0  
## 7 survyr 0  
## 8 \_st 0  
## 9 \_d 0  
## 10 \_t 0  
## 11 \_t0 0  
## 12 ageyr 0  
## 13 agecat 0  
## 14 agegr\_2 0  
## 15 agegr\_3 0  
## 16 hstage2 0  
## 17 hstage3 0  
## 18 hstage4 0

##  
  
  
library(tableone)

## Warning: package 'tableone' was built under R version 4.3.2

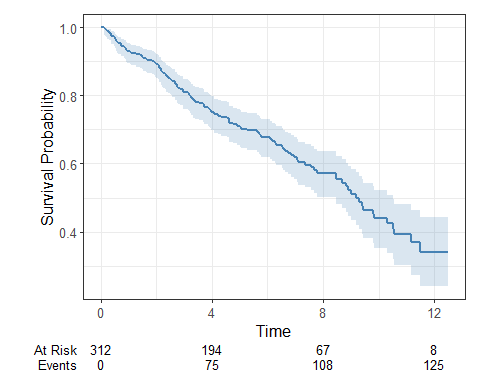
## specify variables  
variables <- c("sex","bil","survyr","death","ageyr","histo","agecat")  
  
## create the table  
table\_1 <- CreateTableOne(vars=variables,strata="drug",data = pbcData,factorVars =   
 c("sex","histo","agecat","death"))  
  
  
## print variables  
print(table\_1,nonnormal = c("bil","survyr","ageyr"))

## Stratified by drug  
## 0 1 p   
## n 154 158   
## sex = Male (%) 15 ( 9.7) 21 (13.3) 0.421  
## bil (median [IQR]) 1.30 [0.72, 3.60] 1.40 [0.80, 3.20] 0.842  
## survyr (median [IQR]) 4.96 [3.16, 7.59] 5.19 [3.37, 7.21] 0.827  
## death = 1 (%) 60 (39.0) 65 (41.1) 0.782  
## ageyr (median [IQR]) 48.14 [41.46, 55.84] 51.97 [43.01, 58.95] 0.020  
## histo (%) 0.201  
## 1 4 ( 2.6) 12 ( 7.6)   
## 2 32 (20.8) 35 (22.2)   
## 3 64 (41.6) 56 (35.4)   
## 4 54 (35.1) 55 (34.8)   
## agecat (%) 0.155  
## < 45 yrs 58 (37.7) 48 (30.4)   
## >= 55 yrs 44 (28.6) 61 (38.6)   
## 45 - 55 yrs 52 (33.8) 49 (31.0)   
## Stratified by drug  
## test   
## n   
## sex = Male (%)   
## bil (median [IQR]) nonnorm  
## survyr (median [IQR]) nonnorm  
## death = 1 (%)   
## ageyr (median [IQR]) nonnorm  
## histo (%)   
## 1   
## 2   
## 3   
## 4   
## agecat (%)   
## < 45 yrs   
## >= 55 yrs   
## 45 - 55 yrs

## survival analysis  
  
## pbcData$SurvObj <- with(pbcData, Surv(survyr, death == 1))  
  
 pbcData <- pbcData |> mutate(survobject=Surv(survyr,death))   
   
 km\_model <- pbcData |>   
 survfit(survobject~ 1,data=\_,conf.type="log-log")  
   
 summary(km\_model)

## Call: survfit(formula = survobject ~ 1, data = pbcData, conf.type = "log-log")  
##   
## time n.risk n.event survival std.err lower 95% CI upper 95% CI  
## 0.112 312 1 0.997 0.00320 0.977 1.000  
## 0.140 311 1 0.994 0.00452 0.975 0.998  
## 0.195 310 1 0.990 0.00552 0.970 0.997  
## 0.211 309 1 0.987 0.00637 0.966 0.995  
## 0.301 308 1 0.984 0.00711 0.962 0.993  
## 0.356 307 1 0.981 0.00778 0.958 0.991  
## 0.359 306 1 0.978 0.00838 0.954 0.989  
## 0.384 305 1 0.974 0.00895 0.949 0.987  
## 0.490 304 1 0.971 0.00948 0.945 0.985  
## 0.510 303 1 0.968 0.00997 0.941 0.983  
## 0.523 302 1 0.965 0.01044 0.937 0.980  
## 0.542 301 1 0.962 0.01089 0.933 0.978  
## 0.567 300 1 0.958 0.01131 0.929 0.976  
## 0.592 299 1 0.955 0.01172 0.925 0.973  
## 0.611 298 1 0.952 0.01211 0.922 0.971  
## 0.723 297 2 0.946 0.01285 0.914 0.966  
## 0.833 295 1 0.942 0.01320 0.910 0.963  
## 0.879 294 1 0.939 0.01354 0.906 0.961  
## 0.893 293 1 0.936 0.01387 0.902 0.958  
## 0.915 292 1 0.933 0.01418 0.899 0.956  
## 0.953 291 1 0.929 0.01449 0.895 0.953  
## 1.063 290 1 0.926 0.01479 0.891 0.950  
## 1.096 289 1 0.923 0.01509 0.887 0.948  
## 1.260 288 1 0.920 0.01537 0.884 0.945  
## 1.411 287 1 0.917 0.01565 0.880 0.942  
## 1.504 285 1 0.913 0.01592 0.876 0.940  
## 1.512 284 1 0.910 0.01619 0.873 0.937  
## 1.636 283 1 0.907 0.01644 0.869 0.934  
## 1.674 282 1 0.904 0.01670 0.865 0.932  
## 1.844 281 1 0.901 0.01695 0.862 0.929  
## 1.901 280 1 0.897 0.01719 0.858 0.926  
## 1.940 279 1 0.894 0.01742 0.854 0.924  
## 2.008 277 1 0.891 0.01766 0.851 0.921  
## 2.055 275 1 0.888 0.01789 0.847 0.918  
## 2.088 274 1 0.884 0.01811 0.843 0.915  
## 2.107 273 1 0.881 0.01833 0.840 0.912  
## 2.153 272 1 0.878 0.01855 0.836 0.910  
## 2.164 270 1 0.875 0.01877 0.833 0.907  
## 2.184 269 1 0.871 0.01898 0.829 0.904  
## 2.189 268 1 0.868 0.01918 0.825 0.901  
## 2.258 267 1 0.865 0.01938 0.822 0.898  
## 2.329 264 1 0.862 0.01958 0.818 0.896  
## 2.337 263 1 0.858 0.01978 0.814 0.893  
## 2.353 262 1 0.855 0.01998 0.811 0.890  
## 2.438 260 1 0.852 0.02017 0.807 0.887  
## 2.477 258 1 0.849 0.02036 0.804 0.884  
## 2.548 257 1 0.845 0.02055 0.800 0.881  
## 2.584 255 1 0.842 0.02073 0.796 0.878  
## 2.660 254 1 0.839 0.02091 0.793 0.875  
## 2.668 253 1 0.835 0.02109 0.789 0.872  
## 2.685 252 1 0.832 0.02127 0.785 0.869  
## 2.737 250 1 0.829 0.02144 0.782 0.866  
## 2.740 249 1 0.825 0.02161 0.778 0.863  
## 2.773 248 1 0.822 0.02178 0.775 0.860  
## 2.841 246 1 0.819 0.02194 0.771 0.857  
## 2.951 244 1 0.815 0.02211 0.767 0.854  
## 2.959 243 1 0.812 0.02227 0.764 0.851  
## 2.967 242 1 0.809 0.02243 0.760 0.848  
## 3.156 239 1 0.805 0.02259 0.756 0.845  
## 3.192 237 1 0.802 0.02275 0.753 0.842  
## 3.205 236 1 0.798 0.02291 0.749 0.839  
## 3.263 235 2 0.792 0.02321 0.742 0.833  
## 3.321 233 1 0.788 0.02336 0.738 0.830  
## 3.334 230 1 0.785 0.02350 0.734 0.827  
## 3.384 227 1 0.781 0.02365 0.731 0.824  
## 3.553 222 1 0.778 0.02381 0.727 0.820  
## 3.699 214 1 0.774 0.02397 0.723 0.817  
## 3.715 213 1 0.771 0.02413 0.719 0.814  
## 3.726 212 1 0.767 0.02429 0.715 0.811  
## 3.871 206 1 0.763 0.02446 0.711 0.807  
## 3.910 203 1 0.759 0.02462 0.707 0.804  
## 3.929 201 1 0.756 0.02479 0.703 0.800  
## 3.956 198 1 0.752 0.02496 0.699 0.797  
## 4.074 193 1 0.748 0.02513 0.695 0.793  
## 4.088 192 1 0.744 0.02530 0.690 0.790  
## 4.208 189 1 0.740 0.02547 0.686 0.786  
## 4.318 184 1 0.736 0.02565 0.682 0.783  
## 4.540 178 1 0.732 0.02583 0.677 0.779  
## 4.608 175 1 0.728 0.02602 0.673 0.775  
## 4.630 174 2 0.719 0.02639 0.664 0.767  
## 4.770 169 1 0.715 0.02657 0.659 0.764  
## 4.893 162 1 0.711 0.02677 0.654 0.760  
## 5.005 159 1 0.706 0.02697 0.650 0.755  
## 5.060 156 1 0.702 0.02718 0.645 0.751  
## 5.274 151 1 0.697 0.02739 0.640 0.747  
## 5.630 141 1 0.692 0.02764 0.634 0.743  
## 5.701 140 1 0.687 0.02788 0.629 0.738  
## 5.726 139 1 0.682 0.02812 0.624 0.734  
## 5.767 138 1 0.677 0.02834 0.618 0.729  
## 6.093 127 1 0.672 0.02862 0.612 0.725  
## 6.181 123 1 0.667 0.02890 0.606 0.720  
## 6.268 121 1 0.661 0.02918 0.600 0.715  
## 6.293 119 1 0.655 0.02946 0.594 0.710  
## 6.537 110 1 0.649 0.02979 0.588 0.704  
## 6.575 109 1 0.644 0.03011 0.581 0.699  
## 6.627 108 1 0.638 0.03041 0.575 0.694  
## 6.756 103 1 0.631 0.03074 0.568 0.688  
## 6.858 100 1 0.625 0.03108 0.561 0.683  
## 6.959 96 1 0.619 0.03143 0.554 0.677  
## 7.077 88 1 0.612 0.03185 0.546 0.671  
## 7.118 87 1 0.604 0.03225 0.538 0.664  
## 7.367 80 1 0.597 0.03272 0.530 0.658  
## 7.586 76 1 0.589 0.03322 0.521 0.651  
## 7.660 74 1 0.581 0.03371 0.512 0.644  
## 7.800 71 1 0.573 0.03421 0.503 0.637  
## 8.455 60 1 0.563 0.03495 0.492 0.629  
## 8.466 59 1 0.554 0.03564 0.481 0.620  
## 8.685 53 1 0.543 0.03646 0.469 0.612  
## 8.827 52 1 0.533 0.03723 0.457 0.603  
## 8.888 50 1 0.522 0.03798 0.445 0.594  
## 8.992 48 1 0.511 0.03872 0.433 0.584  
## 9.200 45 1 0.500 0.03949 0.420 0.574  
## 9.301 43 1 0.488 0.04025 0.407 0.564  
## 9.392 41 1 0.476 0.04099 0.394 0.554  
## 9.438 40 1 0.465 0.04166 0.381 0.544  
## 9.792 37 1 0.452 0.04238 0.368 0.533  
## 9.819 34 1 0.439 0.04317 0.353 0.521  
## 10.307 30 1 0.424 0.04414 0.337 0.509  
## 10.518 27 1 0.408 0.04522 0.319 0.495  
## 10.556 25 1 0.392 0.04626 0.302 0.481  
## 11.175 17 1 0.369 0.04895 0.274 0.464  
## 11.482 13 1 0.341 0.05278 0.240 0.444

## plot km curve  
 km\_model |>   
 ggsurvfit(color="steelblue",linewidth=1)+  
 add\_risktable()+  
 add\_confidence\_interval(fill="steelblue")



# estimate survival curves for entire sample  
km.overall <- survfit(survobject ~ 1, data = pbcData,  
 type="kaplan-meier", conf.type="log-log")  
km.overall

## Call: survfit(formula = survobject ~ 1, data = pbcData, type = "kaplan-meier",   
## conf.type = "log-log")  
##   
## n events median 0.95LCL 0.95UCL  
## [1,] 312 125 9.3 8.45 10.5

summary(km.overall)

## Call: survfit(formula = survobject ~ 1, data = pbcData, type = "kaplan-meier",   
## conf.type = "log-log")  
##   
## time n.risk n.event survival std.err lower 95% CI upper 95% CI  
## 0.112 312 1 0.997 0.00320 0.977 1.000  
## 0.140 311 1 0.994 0.00452 0.975 0.998  
## 0.195 310 1 0.990 0.00552 0.970 0.997  
## 0.211 309 1 0.987 0.00637 0.966 0.995  
## 0.301 308 1 0.984 0.00711 0.962 0.993  
## 0.356 307 1 0.981 0.00778 0.958 0.991  
## 0.359 306 1 0.978 0.00838 0.954 0.989  
## 0.384 305 1 0.974 0.00895 0.949 0.987  
## 0.490 304 1 0.971 0.00948 0.945 0.985  
## 0.510 303 1 0.968 0.00997 0.941 0.983  
## 0.523 302 1 0.965 0.01044 0.937 0.980  
## 0.542 301 1 0.962 0.01089 0.933 0.978  
## 0.567 300 1 0.958 0.01131 0.929 0.976  
## 0.592 299 1 0.955 0.01172 0.925 0.973  
## 0.611 298 1 0.952 0.01211 0.922 0.971  
## 0.723 297 2 0.946 0.01285 0.914 0.966  
## 0.833 295 1 0.942 0.01320 0.910 0.963  
## 0.879 294 1 0.939 0.01354 0.906 0.961  
## 0.893 293 1 0.936 0.01387 0.902 0.958  
## 0.915 292 1 0.933 0.01418 0.899 0.956  
## 0.953 291 1 0.929 0.01449 0.895 0.953  
## 1.063 290 1 0.926 0.01479 0.891 0.950  
## 1.096 289 1 0.923 0.01509 0.887 0.948  
## 1.260 288 1 0.920 0.01537 0.884 0.945  
## 1.411 287 1 0.917 0.01565 0.880 0.942  
## 1.504 285 1 0.913 0.01592 0.876 0.940  
## 1.512 284 1 0.910 0.01619 0.873 0.937  
## 1.636 283 1 0.907 0.01644 0.869 0.934  
## 1.674 282 1 0.904 0.01670 0.865 0.932  
## 1.844 281 1 0.901 0.01695 0.862 0.929  
## 1.901 280 1 0.897 0.01719 0.858 0.926  
## 1.940 279 1 0.894 0.01742 0.854 0.924  
## 2.008 277 1 0.891 0.01766 0.851 0.921  
## 2.055 275 1 0.888 0.01789 0.847 0.918  
## 2.088 274 1 0.884 0.01811 0.843 0.915  
## 2.107 273 1 0.881 0.01833 0.840 0.912  
## 2.153 272 1 0.878 0.01855 0.836 0.910  
## 2.164 270 1 0.875 0.01877 0.833 0.907  
## 2.184 269 1 0.871 0.01898 0.829 0.904  
## 2.189 268 1 0.868 0.01918 0.825 0.901  
## 2.258 267 1 0.865 0.01938 0.822 0.898  
## 2.329 264 1 0.862 0.01958 0.818 0.896  
## 2.337 263 1 0.858 0.01978 0.814 0.893  
## 2.353 262 1 0.855 0.01998 0.811 0.890  
## 2.438 260 1 0.852 0.02017 0.807 0.887  
## 2.477 258 1 0.849 0.02036 0.804 0.884  
## 2.548 257 1 0.845 0.02055 0.800 0.881  
## 2.584 255 1 0.842 0.02073 0.796 0.878  
## 2.660 254 1 0.839 0.02091 0.793 0.875  
## 2.668 253 1 0.835 0.02109 0.789 0.872  
## 2.685 252 1 0.832 0.02127 0.785 0.869  
## 2.737 250 1 0.829 0.02144 0.782 0.866  
## 2.740 249 1 0.825 0.02161 0.778 0.863  
## 2.773 248 1 0.822 0.02178 0.775 0.860  
## 2.841 246 1 0.819 0.02194 0.771 0.857  
## 2.951 244 1 0.815 0.02211 0.767 0.854  
## 2.959 243 1 0.812 0.02227 0.764 0.851  
## 2.967 242 1 0.809 0.02243 0.760 0.848  
## 3.156 239 1 0.805 0.02259 0.756 0.845  
## 3.192 237 1 0.802 0.02275 0.753 0.842  
## 3.205 236 1 0.798 0.02291 0.749 0.839  
## 3.263 235 2 0.792 0.02321 0.742 0.833  
## 3.321 233 1 0.788 0.02336 0.738 0.830  
## 3.334 230 1 0.785 0.02350 0.734 0.827  
## 3.384 227 1 0.781 0.02365 0.731 0.824  
## 3.553 222 1 0.778 0.02381 0.727 0.820  
## 3.699 214 1 0.774 0.02397 0.723 0.817  
## 3.715 213 1 0.771 0.02413 0.719 0.814  
## 3.726 212 1 0.767 0.02429 0.715 0.811  
## 3.871 206 1 0.763 0.02446 0.711 0.807  
## 3.910 203 1 0.759 0.02462 0.707 0.804  
## 3.929 201 1 0.756 0.02479 0.703 0.800  
## 3.956 198 1 0.752 0.02496 0.699 0.797  
## 4.074 193 1 0.748 0.02513 0.695 0.793  
## 4.088 192 1 0.744 0.02530 0.690 0.790  
## 4.208 189 1 0.740 0.02547 0.686 0.786  
## 4.318 184 1 0.736 0.02565 0.682 0.783  
## 4.540 178 1 0.732 0.02583 0.677 0.779  
## 4.608 175 1 0.728 0.02602 0.673 0.775  
## 4.630 174 2 0.719 0.02639 0.664 0.767  
## 4.770 169 1 0.715 0.02657 0.659 0.764  
## 4.893 162 1 0.711 0.02677 0.654 0.760  
## 5.005 159 1 0.706 0.02697 0.650 0.755  
## 5.060 156 1 0.702 0.02718 0.645 0.751  
## 5.274 151 1 0.697 0.02739 0.640 0.747  
## 5.630 141 1 0.692 0.02764 0.634 0.743  
## 5.701 140 1 0.687 0.02788 0.629 0.738  
## 5.726 139 1 0.682 0.02812 0.624 0.734  
## 5.767 138 1 0.677 0.02834 0.618 0.729  
## 6.093 127 1 0.672 0.02862 0.612 0.725  
## 6.181 123 1 0.667 0.02890 0.606 0.720  
## 6.268 121 1 0.661 0.02918 0.600 0.715  
## 6.293 119 1 0.655 0.02946 0.594 0.710  
## 6.537 110 1 0.649 0.02979 0.588 0.704  
## 6.575 109 1 0.644 0.03011 0.581 0.699  
## 6.627 108 1 0.638 0.03041 0.575 0.694  
## 6.756 103 1 0.631 0.03074 0.568 0.688  
## 6.858 100 1 0.625 0.03108 0.561 0.683  
## 6.959 96 1 0.619 0.03143 0.554 0.677  
## 7.077 88 1 0.612 0.03185 0.546 0.671  
## 7.118 87 1 0.604 0.03225 0.538 0.664  
## 7.367 80 1 0.597 0.03272 0.530 0.658  
## 7.586 76 1 0.589 0.03322 0.521 0.651  
## 7.660 74 1 0.581 0.03371 0.512 0.644  
## 7.800 71 1 0.573 0.03421 0.503 0.637  
## 8.455 60 1 0.563 0.03495 0.492 0.629  
## 8.466 59 1 0.554 0.03564 0.481 0.620  
## 8.685 53 1 0.543 0.03646 0.469 0.612  
## 8.827 52 1 0.533 0.03723 0.457 0.603  
## 8.888 50 1 0.522 0.03798 0.445 0.594  
## 8.992 48 1 0.511 0.03872 0.433 0.584  
## 9.200 45 1 0.500 0.03949 0.420 0.574  
## 9.301 43 1 0.488 0.04025 0.407 0.564  
## 9.392 41 1 0.476 0.04099 0.394 0.554  
## 9.438 40 1 0.465 0.04166 0.381 0.544  
## 9.792 37 1 0.452 0.04238 0.368 0.533  
## 9.819 34 1 0.439 0.04317 0.353 0.521  
## 10.307 30 1 0.424 0.04414 0.337 0.509  
## 10.518 27 1 0.408 0.04522 0.319 0.495  
## 10.556 25 1 0.392 0.04626 0.302 0.481  
## 11.175 17 1 0.369 0.04895 0.274 0.464  
## 11.482 13 1 0.341 0.05278 0.240 0.444

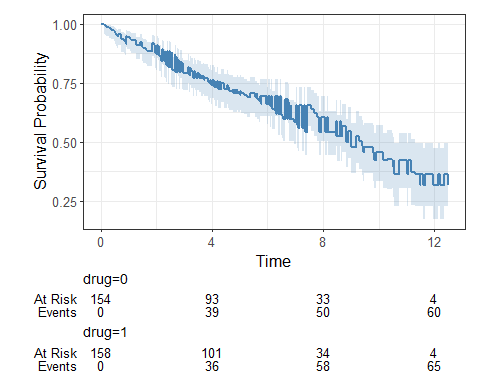
# estimate survival curves for drug group  
km.drug <- survfit(survobject ~ drug, data = pbcData,  
 type="kaplan-meier", conf.type="log-log")  
km.drug

## Call: survfit(formula = survobject ~ drug, data = pbcData, type = "kaplan-meier",   
## conf.type = "log-log")  
##   
## n events median 0.95LCL 0.95UCL  
## drug=0 154 60 9.39 8.47 10.6  
## drug=1 158 65 8.99 6.96 11.5

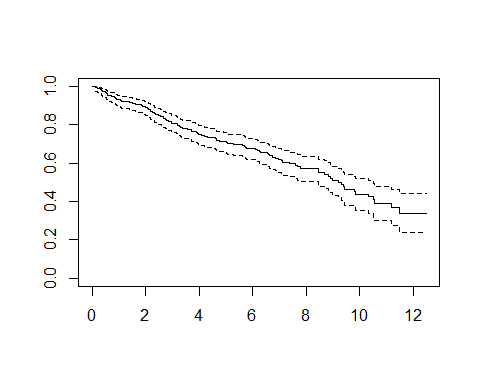
summary(km.drug)

## Call: survfit(formula = survobject ~ drug, data = pbcData, type = "kaplan-meier",   
## conf.type = "log-log")  
##   
## drug=0   
## time n.risk n.event survival std.err lower 95% CI upper 95% CI  
## 0.140 154 1 0.994 0.00647 0.955 0.999  
## 0.211 153 1 0.987 0.00912 0.949 0.997  
## 0.301 152 1 0.981 0.01114 0.941 0.994  
## 0.356 151 1 0.974 0.01282 0.932 0.990  
## 0.510 150 1 0.968 0.01428 0.924 0.986  
## 0.523 149 1 0.961 0.01559 0.915 0.982  
## 0.567 148 1 0.955 0.01679 0.907 0.978  
## 0.592 147 1 0.948 0.01788 0.899 0.974  
## 0.723 146 2 0.935 0.01986 0.883 0.965  
## 0.833 144 1 0.929 0.02075 0.875 0.960  
## 0.879 143 1 0.922 0.02160 0.867 0.955  
## 0.893 142 1 0.916 0.02240 0.859 0.950  
## 1.260 141 1 0.909 0.02317 0.851 0.945  
## 1.504 140 1 0.903 0.02389 0.844 0.940  
## 1.512 139 1 0.896 0.02459 0.836 0.935  
## 1.636 138 1 0.890 0.02525 0.828 0.930  
## 1.674 137 1 0.883 0.02589 0.821 0.925  
## 1.940 136 1 0.877 0.02650 0.813 0.919  
## 2.008 135 1 0.870 0.02709 0.806 0.914  
## 2.107 134 1 0.864 0.02765 0.799 0.909  
## 2.153 133 1 0.857 0.02820 0.791 0.904  
## 2.164 131 1 0.851 0.02873 0.784 0.898  
## 2.184 130 1 0.844 0.02925 0.776 0.893  
## 2.329 128 1 0.837 0.02975 0.769 0.887  
## 2.337 127 1 0.831 0.03024 0.762 0.882  
## 2.353 126 1 0.824 0.03071 0.754 0.876  
## 2.438 125 1 0.818 0.03116 0.747 0.870  
## 2.548 124 1 0.811 0.03160 0.740 0.865  
## 2.584 123 1 0.804 0.03203 0.732 0.859  
## 2.668 122 1 0.798 0.03244 0.725 0.853  
## 2.959 118 1 0.791 0.03286 0.718 0.847  
## 3.192 115 1 0.784 0.03328 0.710 0.841  
## 3.321 114 1 0.777 0.03370 0.703 0.836  
## 3.334 111 1 0.770 0.03411 0.695 0.829  
## 3.715 103 1 0.763 0.03459 0.687 0.823  
## 3.871 101 1 0.755 0.03506 0.678 0.816  
## 3.910 98 1 0.748 0.03554 0.670 0.810  
## 3.956 95 1 0.740 0.03603 0.661 0.803  
## 4.074 93 1 0.732 0.03651 0.652 0.796  
## 4.208 91 1 0.724 0.03698 0.644 0.789  
## 4.893 79 1 0.715 0.03763 0.633 0.781  
## 5.060 76 1 0.705 0.03829 0.623 0.773  
## 5.726 69 1 0.695 0.03908 0.611 0.764  
## 6.627 56 1 0.683 0.04030 0.596 0.754  
## 6.756 53 1 0.670 0.04155 0.581 0.744  
## 6.858 51 1 0.657 0.04276 0.566 0.733  
## 7.586 40 1 0.640 0.04473 0.545 0.720  
## 7.660 38 1 0.623 0.04662 0.525 0.707  
## 7.800 35 1 0.605 0.04857 0.503 0.693  
## 8.466 32 1 0.587 0.05060 0.481 0.678  
## 8.685 29 1 0.566 0.05275 0.457 0.662  
## 8.888 28 1 0.546 0.05460 0.433 0.646  
## 9.200 26 1 0.525 0.05640 0.409 0.628  
## 9.301 24 1 0.503 0.05814 0.385 0.610  
## 9.392 22 1 0.480 0.05983 0.360 0.591  
## 9.438 21 1 0.457 0.06119 0.335 0.572  
## 10.307 15 1 0.427 0.06427 0.300 0.548  
## 10.518 13 1 0.394 0.06719 0.264 0.522  
## 10.556 12 1 0.361 0.06916 0.230 0.494  
##   
## drug=1   
## time n.risk n.event survival std.err lower 95% CI upper 95% CI  
## 0.112 158 1 0.994 0.00631 0.956 0.999  
## 0.195 157 1 0.987 0.00889 0.950 0.997  
## 0.359 156 1 0.981 0.01086 0.942 0.994  
## 0.384 155 1 0.975 0.01250 0.934 0.990  
## 0.490 154 1 0.968 0.01393 0.926 0.987  
## 0.542 153 1 0.962 0.01521 0.917 0.983  
## 0.611 152 1 0.956 0.01637 0.909 0.979  
## 0.915 151 1 0.949 0.01744 0.901 0.974  
## 0.953 150 1 0.943 0.01844 0.893 0.970  
## 1.063 149 1 0.937 0.01937 0.886 0.965  
## 1.096 148 1 0.930 0.02025 0.878 0.961  
## 1.411 147 1 0.924 0.02108 0.870 0.956  
## 1.844 145 1 0.918 0.02187 0.862 0.951  
## 1.901 144 1 0.911 0.02263 0.855 0.946  
## 2.055 141 1 0.905 0.02337 0.847 0.942  
## 2.088 140 1 0.898 0.02408 0.839 0.936  
## 2.189 139 1 0.892 0.02476 0.832 0.931  
## 2.258 138 1 0.885 0.02541 0.824 0.926  
## 2.477 134 1 0.879 0.02607 0.817 0.921  
## 2.660 132 1 0.872 0.02671 0.809 0.916  
## 2.685 131 1 0.866 0.02732 0.801 0.910  
## 2.737 130 1 0.859 0.02791 0.794 0.905  
## 2.740 129 1 0.852 0.02848 0.786 0.899  
## 2.773 128 1 0.846 0.02902 0.778 0.894  
## 2.841 127 1 0.839 0.02955 0.771 0.888  
## 2.951 126 1 0.832 0.03005 0.763 0.883  
## 2.967 125 1 0.826 0.03054 0.756 0.877  
## 3.156 124 1 0.819 0.03101 0.749 0.871  
## 3.205 122 1 0.812 0.03148 0.741 0.866  
## 3.263 121 2 0.799 0.03236 0.726 0.854  
## 3.384 117 1 0.792 0.03279 0.719 0.848  
## 3.553 114 1 0.785 0.03323 0.711 0.842  
## 3.699 111 1 0.778 0.03368 0.703 0.836  
## 3.726 110 1 0.771 0.03411 0.695 0.830  
## 3.929 105 1 0.764 0.03456 0.687 0.823  
## 4.088 100 1 0.756 0.03505 0.679 0.817  
## 4.318 97 1 0.748 0.03554 0.670 0.810  
## 4.540 93 1 0.740 0.03606 0.661 0.803  
## 4.608 92 1 0.732 0.03655 0.653 0.796  
## 4.630 91 2 0.716 0.03748 0.635 0.782  
## 4.770 87 1 0.708 0.03794 0.626 0.775  
## 5.005 82 1 0.699 0.03845 0.616 0.767  
## 5.274 78 1 0.690 0.03899 0.607 0.759  
## 5.630 72 1 0.681 0.03960 0.596 0.751  
## 5.701 71 1 0.671 0.04019 0.585 0.743  
## 5.767 70 1 0.661 0.04074 0.575 0.734  
## 6.093 65 1 0.651 0.04137 0.564 0.725  
## 6.181 63 1 0.641 0.04198 0.552 0.716  
## 6.268 61 1 0.630 0.04259 0.541 0.707  
## 6.293 60 1 0.620 0.04315 0.529 0.698  
## 6.537 54 1 0.608 0.04385 0.517 0.688  
## 6.575 53 1 0.597 0.04450 0.504 0.678  
## 6.959 47 1 0.584 0.04533 0.490 0.667  
## 7.077 42 1 0.570 0.04634 0.474 0.655  
## 7.118 41 1 0.556 0.04725 0.459 0.643  
## 7.367 38 1 0.542 0.04822 0.443 0.631  
## 8.455 28 1 0.522 0.05023 0.420 0.615  
## 8.827 24 1 0.501 0.05264 0.394 0.598  
## 8.992 22 1 0.478 0.05495 0.367 0.580  
## 9.792 18 1 0.451 0.05795 0.336 0.560  
## 9.819 17 1 0.425 0.06032 0.306 0.539  
## 11.175 8 1 0.372 0.07247 0.233 0.510  
## 11.482 7 1 0.319 0.07922 0.173 0.474

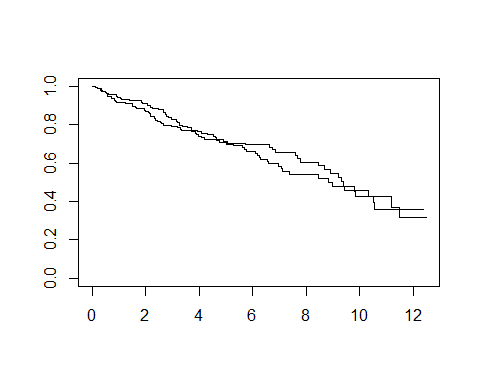
km.drug |>   
 ggsurvfit(color="steelblue",linewidth=1)+  
 add\_risktable()+  
 add\_confidence\_interval(fill="steelblue")



# plot km curves  
plot(km.overall)



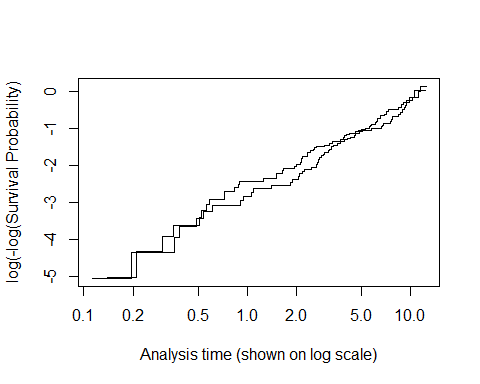
plot(km.drug)



# log rank test for equality of survivor functions  
survdiff(survobject ~ drug, data=pbcData)

## Call:  
## survdiff(formula = survobject ~ drug, data = pbcData)  
##   
## N Observed Expected (O-E)^2/E (O-E)^2/V  
## drug=0 154 60 61.8 0.0513 0.102  
## drug=1 158 65 63.2 0.0502 0.102  
##   
## Chisq= 0.1 on 1 degrees of freedom, p= 0.7

# complimentary log-log plot  
plot(km.drug, fun="cloglog", ylab="log(-log(Survival Probability)",  
 xlab="Analysis time (shown on log scale)")



model1 = coxph(survobject~ drug, data = pbcData)  
summary(model1)

## Call:  
## coxph(formula = survobject ~ drug, data = pbcData)  
##   
## n= 312, number of events= 125   
##   
## coef exp(coef) se(coef) z Pr(>|z|)  
## drug 0.05722 1.05889 0.17916 0.319 0.749  
##   
## exp(coef) exp(-coef) lower .95 upper .95  
## drug 1.059 0.9444 0.7453 1.504  
##   
## Concordance= 0.499 (se = 0.025 )  
## Likelihood ratio test= 0.1 on 1 df, p=0.7  
## Wald test = 0.1 on 1 df, p=0.7  
## Score (logrank) test = 0.1 on 1 df, p=0.7

model2 = coxph(survobject ~ sex + bil + as.factor(histo), data = pbcData)  
summary(model2)

## Call:  
## coxph(formula = survobject ~ sex + bil + as.factor(histo), data = pbcData)  
##   
## n= 312, number of events= 125   
##   
## coef exp(coef) se(coef) z Pr(>|z|)   
## sexMale 0.64275 1.90171 0.23926 2.686 0.00722 \*\*   
## bil 0.15149 1.16357 0.01424 10.637 < 2e-16 \*\*\*  
## as.factor(histo)2 1.64339 5.17269 1.03376 1.590 0.11190   
## as.factor(histo)3 2.03122 7.62340 1.01631 1.999 0.04565 \*   
## as.factor(histo)4 2.90689 18.29988 1.01216 2.872 0.00408 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## exp(coef) exp(-coef) lower .95 upper .95  
## sexMale 1.902 0.52584 1.190 3.040  
## bil 1.164 0.85943 1.132 1.197  
## as.factor(histo)2 5.173 0.19332 0.682 39.233  
## as.factor(histo)3 7.623 0.13118 1.040 55.877  
## as.factor(histo)4 18.300 0.05465 2.517 133.045  
##   
## Concordance= 0.812 (se = 0.017 )  
## Likelihood ratio test= 133.2 on 5 df, p=<2e-16  
## Wald test = 149.2 on 5 df, p=<2e-16  
## Score (logrank) test = 218.8 on 5 df, p=<2e-16