Kaplan Meier and Wilcoxon

Isabella Chittumuri

3/1/2021

To answer the questions below, you will need to use a computer program (from SAS, Stata, SPSS, R or any other package you are familiar with) that computes and plots KM curves and computes the log–rank test.

```
setwd("~/Documents/Hunter College/Spring 2021/Stat 755/HW")
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.0     v dplyr   1.0.5
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(survival)
library(survminer)
```

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

1. For the vets.dat data set described in the presentation:

```
# get data
data("veteran")
```

```
## Warning in data("veteran"): data set 'veteran' not found
```

```
?veteran
```

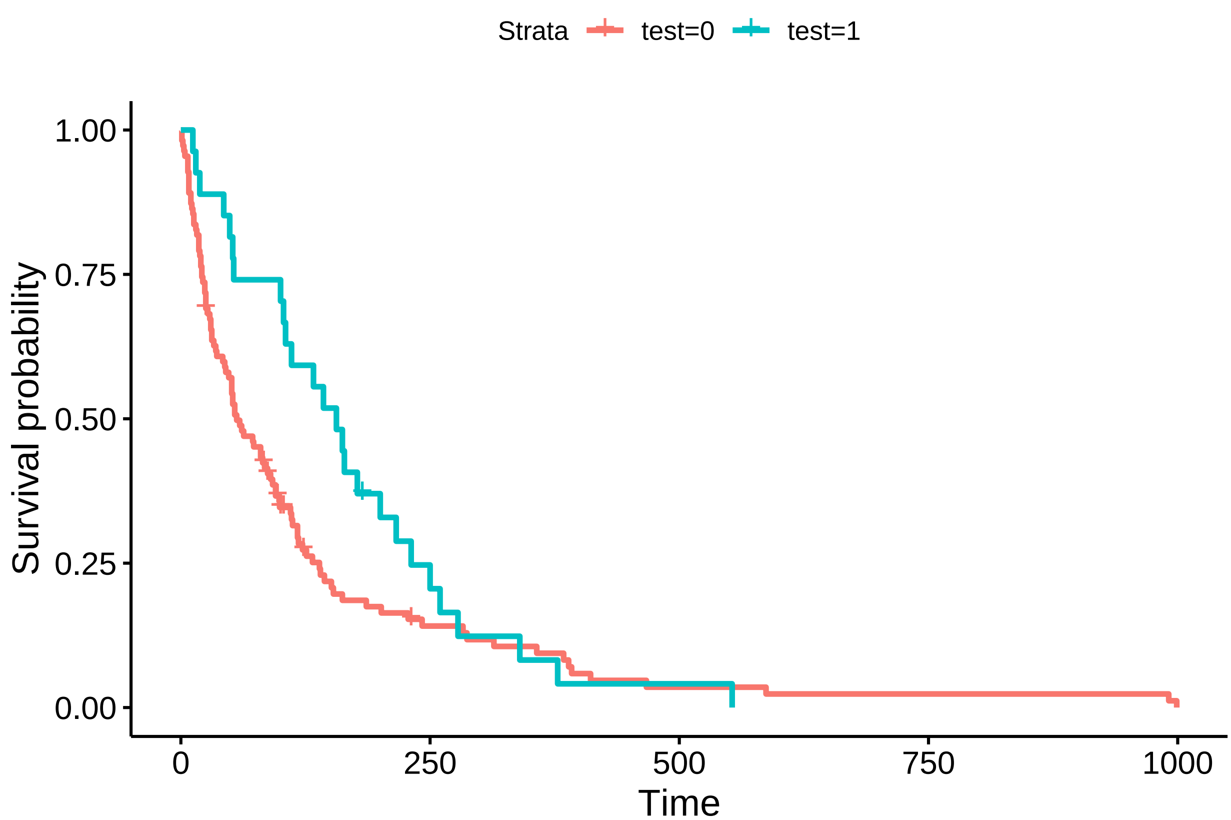
a. Obtain KM plots for the two categories of the variable cell type 1 (1 = large, 0 = other). Comment on how the two curves compare with each other. Carry out the log–rank, and draw conclusions from the test(s).

```
# change celltype categories where 1 = large, 0 = other
veteran$test <- ifelse(veteran$celltype == "large", 1,0)
head(veteran)
```

```
##   trt celltype time status karno diagtime age prior test
## 1   1 squamous   72      1    60        7  69     0    0
## 2   1 squamous  411      1    70        5  64    10    0
## 3   1 squamous  228      1    60        3  38     0    0
## 4   1 squamous  126      1    60        9  63    10    0
## 5   1 squamous  118      1    70       11  65    10    0
## 6   1 squamous   10      1    20        5  49     0    0
```

We created a new column called test where 0 equals celltype:squamous, smallcell, and adeno (aka:other) and 1 equals celltype:large.

```r
# Kaplan-Meier plot
fit <- survfit(Surv(time, status) ~ test, data = veteran)
ggsurvplot(fit)
```



Both the celltype:large curve (blue) and celltype:other curve (red) have similar survival probability slopes. However, celltype:large seems to have slightly higher survival probability over time than celltype:other.

Note celltype:large ends around time 500 whereas celltype:other continuous until time 1000. This may be due to that fact that celltype:other has more observed values.

```r
# log rank test for difference in survival
surv_diff <- survdiff(Surv(time, status) ~ test, data = veteran)
surv_diff
```

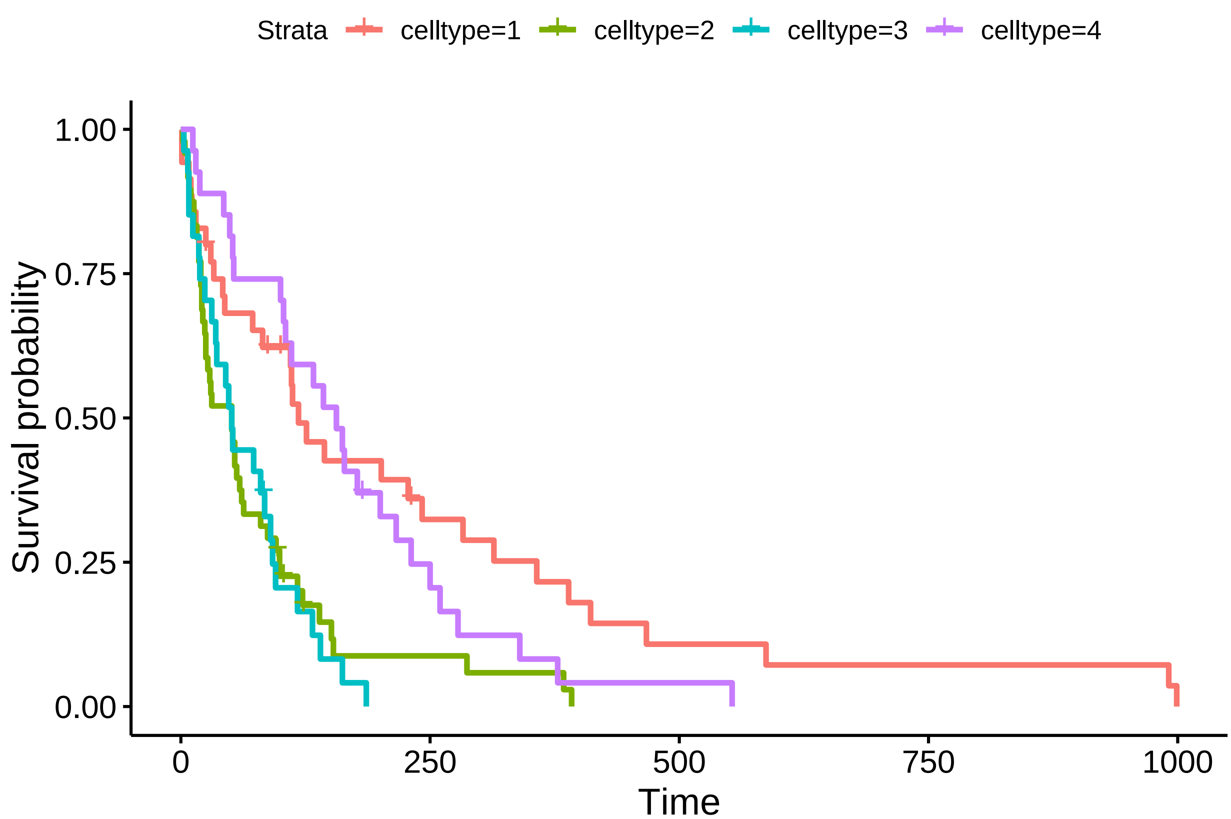
```
## Call:
```

```
## survdiff(formula = Surv(time, status) ~ test, data = veteran)
##
##          N Observed Expected (O-E)^2/E (O-E)^2/V
## test=0 110      102     93.5     0.782      3.02
## test=1  27       26     34.5     2.116      3.02
##
##  Chisq= 3  on 1 degrees of freedom, p= 0.08
```

The log-rank test compares the observed number of events to the expected number of events in each group if the null hypothesis were true (i.e, if the survival curves were identical). In this case, since the resulting p-value of 0.08 is greater than alpha level 0.05, we fail to reject the null hypothesis which means that the survival curves are identical.

b. Obtain KM plots for the four categories of cell type large, adeno, small, and squamous. Note that you will need to recode the data to define a single variable which numerically distinguishes the four categories (e.g., 1 = large, 2 = adeno, etc.). As in part a, compare the four KM curves. Also, carry out the log–rank for the equality of the four curves and draw conclusions.

```r
# Kaplan-Meier plot
fit <- survfit(Surv(time, status) ~ celltype, data = veteran)
ggsurvplot(fit)
```



Squamous (celltype 1, red curve), smallcell (celltype 2, green curve), adeno (celltype 3, blue curve), large (celltype 4, purple curve)

After separating each celltype into its own group, we see that squamous and large have similar curves while small cell and adeno also have similar curves. The KM plot shows that as time increases, survival probability for large and squamous cells is larger than that of small and adeno cells. This means that the former has a longer period of survival than the latter.

```
# log rank test for difference in survival
surv_diff <- survdiff(Surv(time, status) ~ celltype, data = veteran)
surv_diff
```

```
## Call:
## survdiff(formula = Surv(time, status) ~ celltype, data = veteran)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## celltype=1 35       31     47.7      5.82     10.53
## celltype=2 48       45     30.1      7.37     10.20
## celltype=3 27       26     15.7      6.77      8.19
## celltype=4 27       26     34.5      2.12      3.02
##
##  Chisq= 25.4  on 3 degrees of freedom, p= 1e-05
```

After computing the log-rank test, we get a p-value of 1e-05 which is significantly lower than our alpha level of 0.05. This means that we can reject the null hypothesis that the four curves are similar.

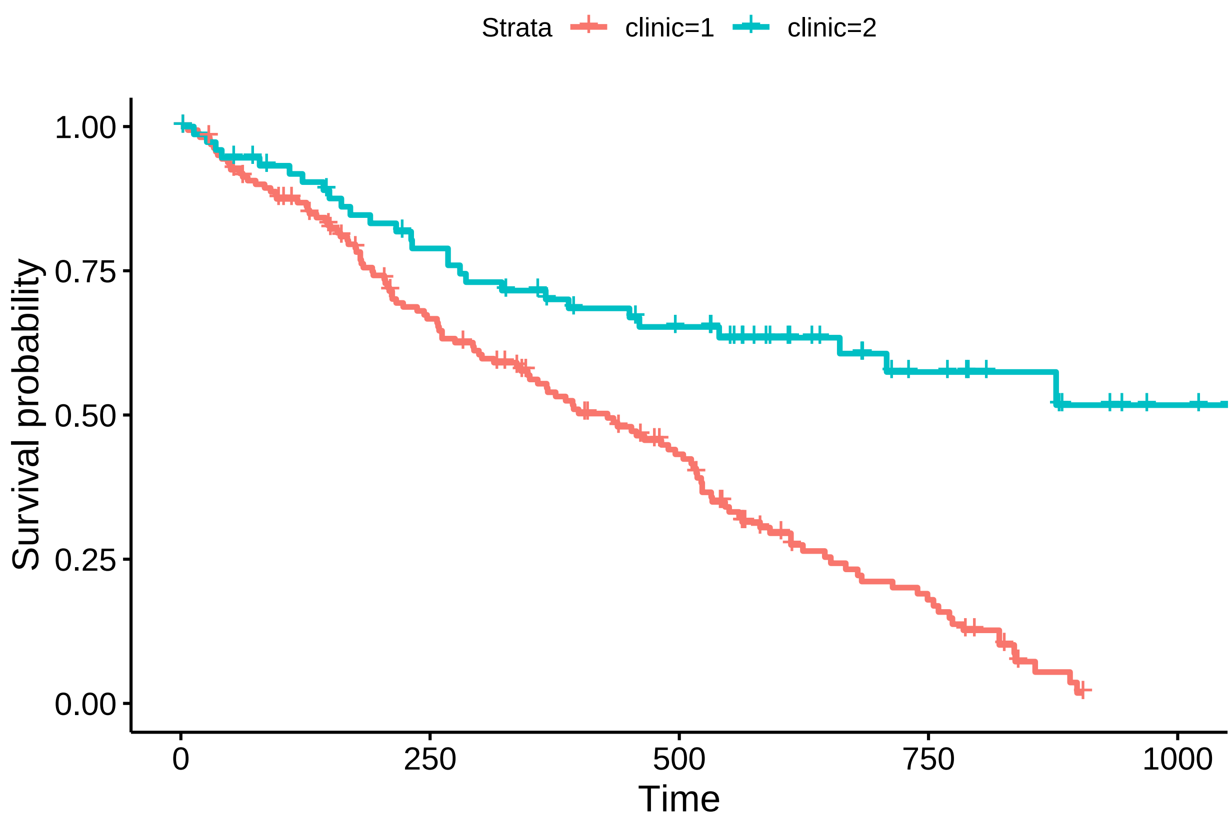
2. The following questions consider a data set from a study by Caplehorn et al. ("Methadone Dosage and Retention of Patients in Maintenance Treatment," Med. J. Aust., 1991). These data comprise the times in days spent by heroin addicts from entry to departure from one of two methadone clinics. There are two fur- ther covariates, namely, prison record and methadone dose, believed to affect the survival times. The data set name is addicts.dat. A listing of the variables is given below:

```
# get data
df <- read.table("addicts.dat")
addicts <- rename(df, subject_ID = V1, clinic = V2, status = V3, time = V4, prison_record = V5, methadon
head(addicts)
```

```
##   subject_ID clinic status time prison_record methadone_dose
## 1          1      1      1  428             0             50
## 2          2      1      1  275             1             55
## 3          3      1      1  262             0             55
## 4          4      1      1  183             0             30
## 5          5      1      1  259             1             65
## 6          6      1      1  714             0             55
```

a. Compute and plot the KM plots for the two categories of the "clinic" variable and comment on the extent to which they differ.

```
# Kaplan-Meier plot
fit <- survfit(Surv(time, status) ~ clinic, data = addicts)
ggsurvplot(fit)
```



From the KM plot, we see that as time increases the survival rate for clinic 1 is much higher than that of clinic 2.

b. A printout of the log–rank and Wilcoxon tests (using Stata) is provided below. What are your conclusions from this printout?

```r
# log rank test for difference in survival
surv_diff <- survdiff(Surv(time, status) ~ clinic, data = addicts)
surv_diff
```

```
## Call:
## survdiff(formula = Surv(time, status) ~ clinic, data = addicts)
##
##             N Observed Expected (O-E)^2/E (O-E)^2/V
## clinic=1 163      122     90.9      10.6      27.9
## clinic=2  75       28     59.1      16.4      27.9
##
##  Chisq= 27.9  on 1 degrees of freedom, p= 1e-07
```

Between group 1 and 2, their difference in observed values is 94 while their difference in expected values is 31.82. This suggests that the two curves have different hazard rates.

Since is the Wilcoxon test puts a higher weight on the earlier values of the model, its test statistic should be higher than that of the log-rank. However, in this case the Wilcoxon test statistic is lower than the log-rank. From this observation, we know that the difference in curves is more evident in the later stage.

The p-value for Wilcoxon is slightly higher than that of log-rank, suggesting that the log-rank is more significant. However, the p-value for both tests are lower than the alpha level of 0.05, which mean that we can reject null hypothesis that the two survival curves are similar.

c. Compute and evaluate KM curves and the log–rank test for comparing suitably chosen categories of the variable "Methadone dose." Explain how you deter- mined the categories for this variable.

```
# get mean
summary(addicts)
```

```
##    subject_ID         clinic          status            time       
##  Min.   :  1.00   Min.   :1.000   Min.   :0.0000   Min.   :   2.0  
##  1st Qu.: 65.25   1st Qu.:1.000   1st Qu.:0.0000   1st Qu.: 171.2  
##  Median :131.50   Median :1.000   Median :1.0000   Median : 367.5  
##  Mean   :134.13   Mean   :1.315   Mean   :0.6303   Mean   : 402.6  
##  3rd Qu.:205.75   3rd Qu.:2.000   3rd Qu.:1.0000   3rd Qu.: 585.5  
##  Max.   :266.00   Max.   :2.000   Max.   :1.0000   Max.   :1076.0  
##  prison_record    methadone_dose 
##  Min.   :0.0000   Min.   : 20.0  
##  1st Qu.:0.0000   1st Qu.: 50.0  
##  Median :0.0000   Median : 60.0  
##  Mean   :0.4664   Mean   : 60.4  
##  3rd Qu.:1.0000   3rd Qu.: 70.0  
##  Max.   :1.0000   Max.   :110.0
```

```
# create groups of 3
# addicts <- addicts %>% mutate(group = ifelse(methadone_dose <= 50, 1, ifelse((methadone_dose > 50 & m

# create groups of 2
addicts <- addicts %>% mutate(group = ifelse(methadone_dose <= 60, 1, 2))
addicts
```

```
##     subject_ID clinic status time prison_record methadone_dose group
## 1            1      1      1  428             0             50     1
## 2            2      1      1  275             1             55     1
## 3            3      1      1  262             0             55     1
## 4            4      1      1  183             0             30     1
## 5            5      1      1  259             1             65     2
## 6            6      1      1  714             0             55     1
## 7            7      1      1  438             1             65     2
## 8            8      1      0  796             1             60     1
## 9            9      1      1  892             0             50     1
## 10          10      1      1  393             1             65     2
## 11          11      1      0  161             1             80     2
## 12          12      1      1  836             1             60     1
## 13          13      1      1  523             0             55     1
## 14          14      1      1  612             0             70     2
## 15          15      1      1  212             1             60     1
## 16          16      1      1  399             1             60     1
## 17          17      1      1  771             1             75     2
## 18          18      1      1  514             1             80     2
## 19          19      1      1  512             0             80     2
## 20          21      1      1  624             1             80     2
```

```
## 21            22        1        1   209                 1                 60       1
## 22            23        1        1   341                 1                 60       1
## 23            24        1        1   299                 0                 55       1
## 24            25        1        0   826                 0                 80       2
## 25            26        1        1   262                 1                 65       2
## 26            27        1        0   566                 1                 45       1
## 27            28        1        1   368                 1                 55       1
## 28            30        1        1   302                 1                 50       1
## 29            31        1        0   602                 0                 60       1
## 30            32        1        1   652                 0                 80       2
## 31            33        1        1   293                 0                 65       2
## 32            34        1        0   564                 0                 60       1
## 33            36        1        1   394                 1                 55       1
## 34            37        1        1   755                 1                 65       2
## 35            38        1        1   591                 0                 55       1
## 36            39        1        0   787                 0                 80       2
## 37            40        1        1   739                 0                 60       1
## 38            41        1        1   550                 1                 60       1
## 39            42        1        1   837                 0                 60       1
## 40            43        1        1   612                 0                 65       2
## 41            44        1        0   581                 0                 70       2
## 42            45        1        1   523                 0                 60       1
## 43            46        1        1   504                 1                 60       1
## 44            48        1        1   785                 1                 80       2
## 45            49        1        1   774                 1                 65       2
## 46            50        1        1   560                 0                 65       2
## 47            51        1        1   160                 0                 35       1
## 48            52        1        1   482                 0                 30       1
## 49            53        1        1   518                 0                 65       2
## 50            54        1        1   683                 0                 50       1
## 51            55        1        1   147                 0                 65       2
## 52            57        1        1   563                 1                 70       2
## 53            58        1        1   646                 1                 60       1
## 54            59        1        1   899                 0                 60       1
## 55            60        1        1   857                 0                 60       1
## 56            61        1        1   180                 1                 70       2
## 57            62        1        1   452                 0                 60       1
## 58            63        1        1   760                 0                 60       1
## 59            64        1        1   496                 0                 65       2
## 60            65        1        1   258                 1                 40       1
## 61            66        1        1   181                 1                 60       1
## 62            67        1        1   386                 0                 60       1
## 63            68        1        0   439                 0                 80       2
## 64            69        1        0   563                 0                 75       2
## 65            70        1        1   337                 0                 65       2
## 66            71        1        0   613                 1                 60       1
## 67            72        1        1   192                 1                 80       2
## 68            73        1        0   405                 0                 80       2
## 69            74        1        1   667                 0                 50       1
## 70            75        1        0   905                 0                 80       2
## 71            76        1        1   247                 0                 70       2
## 72            77        1        1   821                 0                 80       2
## 73            78        1        1   821                 1                 75       2
## 74            79        1        0   517                 0                 45       1
```

```
## 75          80   1   0  346           1         60   1
## 76          81   1   1  294           0         65   2
## 77          82   1   1  244           1         60   1
## 78          83   1   1   95           1         60   1
## 79          84   1   1  376           1         55   1
## 80          85   1   1  212           0         40   1
## 81          86   1   1   96           0         70   2
## 82          87   1   1  532           0         80   2
## 83          88   1   1  522           1         70   2
## 84          89   1   1  679           0         35   1
## 85          90   1   0  408           0         50   1
## 86          91   1   0  840           0         80   2
## 87          92   1   0  148           1         65   2
## 88          93   1   1  168           0         65   2
## 89          94   1   1  489           0         80   2
## 90          95   1   0  541           0         80   2
## 91          96   1   1  205           0         50   1
## 92          97   1   0  475           1         75   2
## 93          98   1   1  237           0         45   1
## 94          99   1   1  517           0         70   2
## 95         100   1   1  749           0         70   2
## 96         101   1   1  150           1         80   2
## 97         102   1   1  465           0         65   2
## 98         103   2   1  708           1         60   1
## 99         104   2   0  713           0         50   1
## 100        105   2   0  146           0         50   1
## 101        106   2   1  450           0         55   1
## 102        109   2   0  555           0         80   2
## 103        110   2   1  460           0         50   1
## 104        111   2   0   53           1         60   1
## 105        113   2   1  122           1         60   1
## 106        114   2   1   35           1         40   1
## 107        118   2   0  532           0         70   2
## 108        119   2   0  684           0         65   2
## 109        120   2   0  769           1         70   2
## 110        121   2   0  591           0         70   2
## 111        122   2   0  769           1         40   1
## 112        123   2   0  609           1        100   2
## 113        124   2   0  932           1         80   2
## 114        125   2   0  932           1         80   2
## 115        126   2   0  587           0        110   2
## 116        127   2   1   26           0         40   1
## 117        128   2   0   72           1         40   1
## 118        129   2   0  641           0         70   2
## 119        131   2   0  367           0         70   2
## 120        132   2   0  633           0         70   2
## 121        133   2   1  661           0         40   1
## 122        134   2   1  232           1         70   2
## 123        135   2   1   13           1         60   1
## 124        137   2   0  563           0         70   2
## 125        138   2   0  969           0         80   2
## 126        143   2   0 1052           0         80   2
## 127        144   2   0  944           1         80   2
## 128        145   2   0  881           0         80   2
```

##	129	146	2	1	190	1	50	1
##	130	148	2	1	79	0	40	1
##	131	149	2	0	884	1	50	1
##	132	150	2	1	170	0	40	1
##	133	153	2	1	286	0	45	1
##	134	156	2	0	358	0	60	1
##	135	158	2	0	326	1	60	1
##	136	159	2	0	769	1	40	1
##	137	160	2	1	161	0	40	1
##	138	161	2	0	564	1	80	2
##	139	162	2	1	268	1	70	2
##	140	163	2	0	611	1	40	1
##	141	164	2	1	322	0	55	1
##	142	165	2	0	1076	1	80	2
##	143	166	2	0	2	1	40	1
##	144	168	2	0	788	0	70	2
##	145	169	2	0	575	0	80	2
##	146	170	2	1	109	1	70	2
##	147	171	2	0	730	1	80	2
##	148	172	2	0	790	0	90	2
##	149	173	2	0	456	1	70	2
##	150	175	2	1	231	1	60	1
##	151	176	2	1	143	1	70	2
##	152	177	2	0	86	1	40	1
##	153	178	2	0	1021	0	80	2
##	154	179	2	0	684	1	80	2
##	155	180	2	1	878	1	60	1
##	156	181	2	1	216	0	100	2
##	157	182	2	0	808	0	60	1
##	158	183	2	1	268	1	40	1
##	159	184	2	0	222	0	40	1
##	160	186	2	0	683	0	100	2
##	161	187	2	0	496	0	40	1
##	162	188	2	1	389	0	55	1
##	163	189	1	1	126	1	75	2
##	164	190	1	1	17	1	40	1
##	165	192	1	1	350	0	60	1
##	166	193	2	0	531	1	65	2
##	167	194	1	0	317	1	50	1
##	168	195	1	0	461	1	75	2
##	169	196	1	1	37	0	60	1
##	170	197	1	1	167	1	55	1
##	171	198	1	1	358	0	45	1
##	172	199	1	1	49	0	60	1
##	173	200	1	1	457	1	40	1
##	174	201	1	1	127	0	20	1
##	175	202	1	1	7	1	40	1
##	176	203	1	1	29	1	60	1
##	177	204	1	1	62	0	40	1
##	178	205	1	0	150	1	60	1
##	179	206	1	1	223	1	40	1
##	180	207	1	0	129	1	40	1
##	181	208	1	0	204	1	65	2
##	182	209	1	1	129	1	50	1

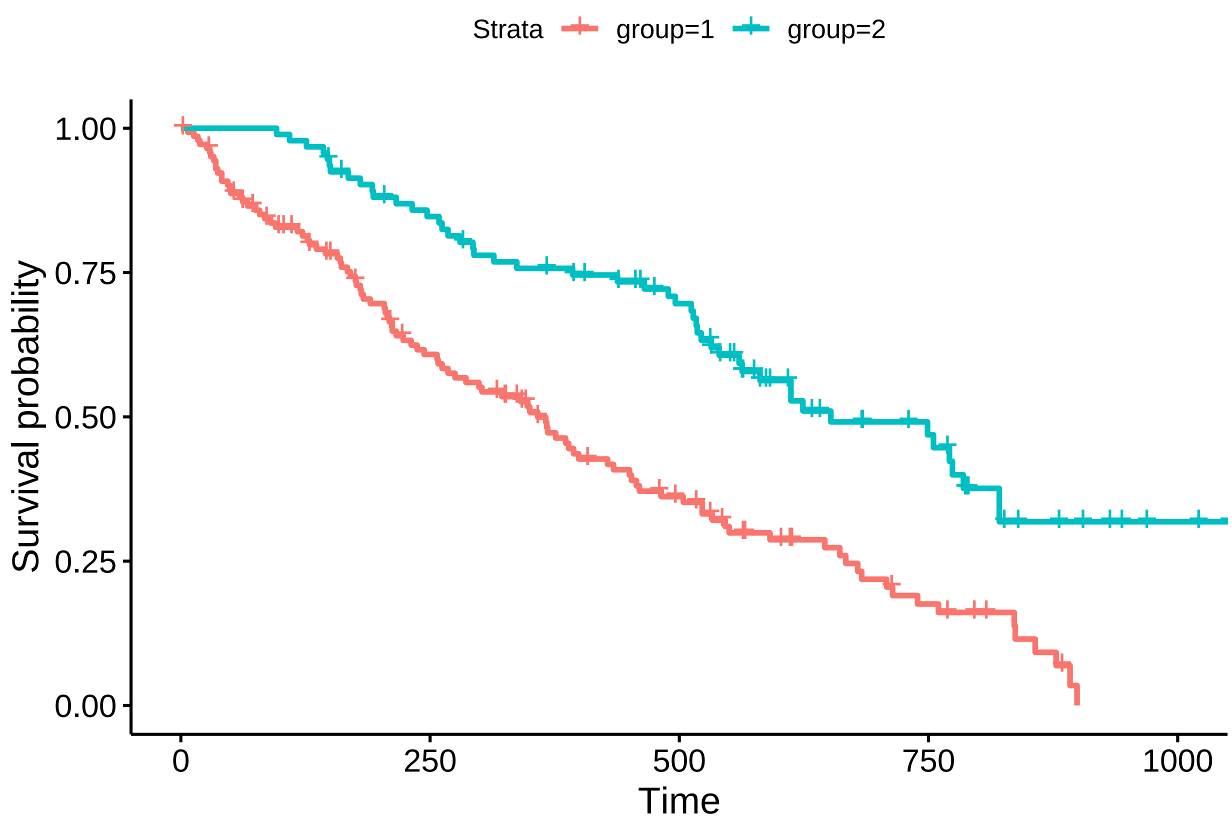
```
## 183          210      1      1  581                 0                 65       2
## 184          211      1      1  176                 0                 55       1
## 185          212      1      1   30                 0                 60       1
## 186          213      1      1   41                 0                 60       1
## 187          214      1      0  543                 0                 40       1
## 188          215      1      0  210                 1                 50       1
## 189          216      1      1  193                 1                 70       2
## 190          217      1      1  434                 0                 55       1
## 191          218      1      1  367                 0                 45       1
## 192          219      1      1  348                 1                 60       1
## 193          220      1      0   28                 0                 50       1
## 194          221      1      0  337                 0                 40       1
## 195          222      1      0  175                 1                 60       1
## 196          223      2      1  149                 1                 80       2
## 197          224      1      1  546                 1                 50       1
## 198          225      1      1   84                 0                 45       1
## 199          226      1      0  283                 1                 80       2
## 200          227      1      1  533                 0                 55       1
## 201          228      1      1  207                 1                 50       1
## 202          229      1      1  216                 0                 50       1
## 203          230      1      0   28                 0                 50       1
## 204          231      1      1   67                 1                 50       1
## 205          232      1      0   62                 1                 60       1
## 206          233      1      0  111                 0                 55       1
## 207          234      1      1  257                 1                 60       1
## 208          235      1      1  136                 1                 55       1
## 209          236      1      0  342                 0                 60       1
## 210          237      2      1   41                 0                 40       1
## 211          238      2      0  531                 1                 45       1
## 212          239      1      0   98                 0                 40       1
## 213          240      1      1  145                 1                 55       1
## 214          241      1      1   50                 0                 50       1
## 215          242      1      0   53                 0                 50       1
## 216          243      1      0  103                 1                 50       1
## 217          244      1      0    2                 1                 60       1
## 218          245      1      1  157                 1                 60       1
## 219          246      1      1   75                 1                 55       1
## 220          247      1      1   19                 1                 40       1
## 221          248      1      1   35                 0                 60       1
## 222          249      2      0  394                 1                 80       2
## 223          250      1      1  117                 0                 40       1
## 224          251      1      1  175                 1                 60       1
## 225          252      1      1  180                 1                 60       1
## 226          253      1      1  314                 0                 70       2
## 227          254      1      0  480                 0                 50       1
## 228          255      1      0  325                 1                 60       1
## 229          256      2      1  280                 0                 90       2
## 230          257      1      1  204                 0                 50       1
## 231          258      2      1  366                 0                 55       1
## 232          259      2      0  531                 1                 50       1
## 233          260      1      1   59                 1                 45       1
## 234          261      1      1   33                 1                 60       1
## 235          262      2      1  540                 0                 80       2
## 236          263      2      0  551                 0                 65       2
```

```
## 237          264       1       1    90               0              40      1
## 238          266       1       1    47               0              45      1
```

```r
# count groups
addicts %>% count(group)
```

```
##   group   n
## 1     1 145
## 2     2  93
```

```r
# Kaplan-Meier plot
fit <- survfit(Surv(time, status) ~ group, data = addicts)
ggsurvplot(fit)
```



We separated the Methadone dose into 3 groups where group 1 has values within the 1st quantile (<=50), group 2 has values of the 2 quantile (>50 & <=70) and group 3 has values of the 3 quantile (<70).

From the KM plot, we see that group 1 and 2 have similar survival curves with steep slopes while group 3 has a higher survival curve with little steepness in slope. This means that as time increases, methadone dose values of less than 70 have a lower survival rate than that of dose values higher than 70. In conclusion, the higher the methadone dose values, the higher chance of survival.

Notes

0.05 alpha level Chisq test is testing for if the curves are similar. p<0.05 = reject null # want to reject null because it shows that the two curves are different. p>0.05 = fail to reject null

Higher test statistic = more significant Lower p-value = more significant