Lab 5

1. Do the chocolate chip melting activity as a class! Submit your responses to the link on PolyLearn by Lab 5.
2. R The text file `Graduate.txt` folder contains the time (in years) that 1000 students (472 males and 528 females) took to graduate (obtain a bachelor's degree) from college (measured from the time they entered a post-secondary institution, i.e. either a junior college or four year degree granting institution). The `Gender` column contains the gender of each student (`0 = male`, `1 = female`), and `Censor` contains the values of the censoring status variable.

   (a) Save the data to your computer and read the data into R. Here is an example command you can use. *No written response required.*

```
graduate <- read.table("FILE PATH/Graduate.txt", header = TRUE)
```

```
library(survival)
graduate <- read.table("C:/Users/spileggi/Google Drive/STAT 417/Data sets/Graduate.txt",
                       header = TRUE)
head(graduate)
```

```
##     Years Censor Gender
## 1 7.50000      0      0
## 2 3.66667      1      1
## 3 0.58333      0      0
## 4 3.66667      1      1
## 5 3.75000      1      1
## 6 4.25000      0      0
```

   (b) Construct Kaplan-Meier estimates to compare the "survival" experiences among male and female college students (i.e., create the `KM_obj`). Plot the Kaplan-Meier estimates to compare the "survival" experiences in the samples of male and female college students by the method of your choice. Comment on what you observe in the survival curves.

```
KM_obj <- survfit(Surv(Years, Censor) ~ Gender,
                  data = graduate)
```

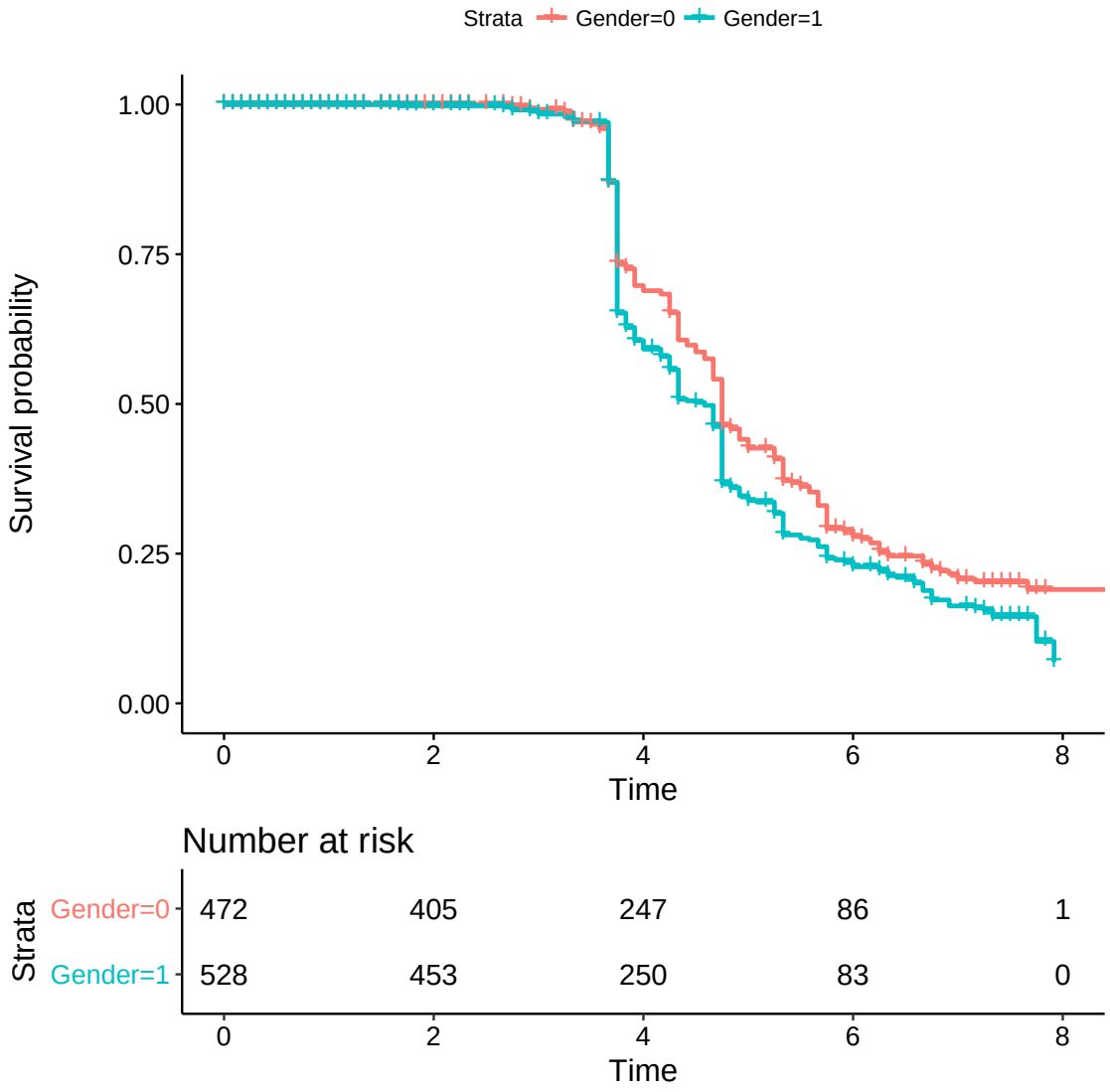
```
library(survminer)
```

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

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

```
## Loading required package: magrittr
```

```
ggsurvplot(KM_obj, data = graduate, risk.table = TRUE)
```



It appears that females tend to graduate sooner than males.

(c) Report the median "survival" time by gender.

```
KM_obj
```

```
## Call: survfit(formula = Surv(Years, Censor) ~ Gender, data = graduate)
##
##             n events median 0.95LCL 0.95UCL
## Gender=0 472    273   4.75    4.67    4.92
## Gender=1 528    341   4.58    4.33    4.75
```

(d) Obtain the Kaplan-Meier estimates of survival probabilities by gender. Report the estimated probabilities that time to graduate exceeds 4 years for both males and females.

```
summary(KM_obj)
```

```
## Call: survfit(formula = Surv(Years, Censor) ~ Gender, data = graduate)
##
```

```
##                Gender=0
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##  2.75    383       1    0.997 0.00261        0.992        1.000
##  2.92    377       2    0.992 0.00454        0.983        1.000
##  3.25    371       2    0.987 0.00589        0.975        0.998
##  3.33    367       6    0.971 0.00873        0.954        0.988
##  3.50    358       1    0.968 0.00912        0.950        0.986
##  3.58    356       3    0.960 0.01018        0.940        0.980
##  3.67    351      33    0.870 0.01757        0.836        0.905
##  3.75    315      49    0.734 0.02314        0.690        0.781
##  3.83    262       3    0.726 0.02338        0.681        0.773
##  3.92    257      10    0.698 0.02411        0.652        0.747
##  4.00    247       3    0.689 0.02431        0.643        0.738
##  4.17    244       2    0.683 0.02444        0.637        0.733
##  4.25    242      11    0.652 0.02506        0.605        0.703
##  4.33    229      16    0.607 0.02577        0.558        0.660
##  4.42    213       3    0.598 0.02587        0.550        0.651
##  4.50    210       4    0.587 0.02600        0.538        0.640
##  4.58    206       4    0.575 0.02611        0.527        0.629
##  4.67    202      12    0.541 0.02636        0.492        0.596
##  4.75    190      27    0.464 0.02645        0.415        0.519
##  4.83    159       2    0.459 0.02643        0.410        0.513
##  4.92    156       6    0.441 0.02638        0.392        0.496
##  5.00    150       5    0.426 0.02631        0.378        0.481
##  5.25    143       6    0.408 0.02620        0.360        0.463
##  5.33    134      12    0.372 0.02589        0.324        0.426
##  5.42    120       1    0.369 0.02586        0.321        0.423
##  5.50    118       2    0.362 0.02580        0.315        0.417
##  5.58    113       3    0.353 0.02570        0.306        0.407
##  5.67    110       7    0.330 0.02543        0.284        0.384
##  5.75    103      12    0.292 0.02477        0.247        0.345
##  5.92     89       1    0.289 0.02471        0.244        0.341
##  6.00     86       3    0.279 0.02452        0.234        0.331
##  6.08     80       1    0.275 0.02446        0.231        0.327
##  6.17     77       2    0.268 0.02434        0.224        0.320
##  6.25     75       4    0.254 0.02407        0.211        0.305
##  6.33     67       2    0.246 0.02394        0.203        0.298
##  6.67     61       3    0.234 0.02376        0.192        0.285
##  6.75     56       2    0.226 0.02363        0.184        0.277
##  6.83     51       1    0.221 0.02358        0.179        0.273
##  6.92     49       1    0.217 0.02353        0.175        0.268
##  7.00     48       2    0.208 0.02340        0.166        0.259
##  7.17     43       1    0.203 0.02335        0.162        0.254
##  7.67     16       1    0.190 0.02509        0.147        0.246
##  8.58      1       1    0.000     NaN           NA           NA
##
##                Gender=1
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##  1.67    473       1    0.998 0.00211       0.9938        1.000
##  2.67    437       1    0.996 0.00311       0.9895        1.000
##  2.75    433       2    0.991 0.00448       0.9823        1.000
##  2.92    428       1    0.989 0.00503       0.9789        0.999
##  3.00    426       2    0.984 0.00598       0.9724        0.996
##  3.25    422       2    0.979 0.00680       0.9661        0.993
##  3.33    420       3    0.972 0.00786       0.9571        0.988
```

```
##  3.58    412       1    0.970 0.00819        0.9541        0.986
##  3.67    410      42    0.871 0.01628        0.8393        0.903
##  3.75    366      92    0.652 0.02320        0.6079        0.699
##  3.83    271      10    0.628 0.02356        0.5832        0.676
##  3.92    260       9    0.606 0.02383        0.5611        0.655
##  4.00    250       6    0.591 0.02399        0.5463        0.640
##  4.17    243       5    0.579 0.02410        0.5339        0.629
##  4.25    236       9    0.557 0.02428        0.5116        0.607
##  4.33    225      20    0.508 0.02452        0.4618        0.558
##  4.42    202       1    0.505 0.02453        0.4593        0.556
##  4.50    201       1    0.503 0.02453        0.4568        0.553
##  4.58    199       2    0.498 0.02455        0.4517        0.548
##  4.67    197      14    0.462 0.02455        0.4165        0.513
##  4.75    181      37    0.368 0.02395        0.3237        0.418
##  4.83    141       3    0.360 0.02386        0.3161        0.410
##  4.92    136       5    0.347 0.02371        0.3032        0.396
##  5.00    131       3    0.339 0.02360        0.2955        0.388
##  5.08    127       1    0.336 0.02357        0.2929        0.386
##  5.25    124       7    0.317 0.02330        0.2746        0.366
##  5.33    115      13    0.281 0.02269        0.2401        0.329
##  5.50    101       2    0.276 0.02258        0.2348        0.324
##  5.58     99       1    0.273 0.02252        0.2322        0.321
##  5.67     98       4    0.262 0.02228        0.2215        0.309
##  5.75     94       7    0.242 0.02181        0.2031        0.289
##  5.83     86       1    0.239 0.02173        0.2004        0.286
##  5.92     85       1    0.237 0.02166        0.1978        0.283
##  6.00     83       3    0.228 0.02143        0.1897        0.274
##  6.25     78       2    0.222 0.02128        0.1842        0.268
##  6.33     74       3    0.213 0.02104        0.1757        0.259
##  6.42     70       1    0.210 0.02096        0.1729        0.256
##  6.58     68       3    0.201 0.02071        0.1642        0.246
##  6.67     64       4    0.188 0.02034        0.1524        0.233
##  6.75     60       5    0.173 0.01982        0.1379        0.216
##  6.92     52       3    0.163 0.01950        0.1286        0.206
##  7.17     48       1    0.159 0.01938        0.1255        0.202
##  7.25     45       1    0.156 0.01927        0.1222        0.199
##  7.33     43       3    0.145 0.01892        0.1122        0.187
##  7.75      7       2    0.104 0.02819        0.0607        0.177
##  7.92      3       1    0.069 0.03386        0.0264        0.181
```

Males: $\hat{S}(4) = 0.689$

Females: $\hat{S}(4) = 0.591$

(e) State the null and alternative hypotheses in words and using symbols to compare the overall survival experiences of males and females.

$H_0 : S_1(t) = S_2(t)$ for all t. The probability of not graduating after any time t is the same for males and females.

$H_a : S_1(t) \neq S_2(t)$ for some t. The probability of not graduating after any time t differs for males and females for some time t.

(f) Perform the log-rank test to compare the "survival" experiences of the population of male and female college students. Be sure to report the value of the test statistic and the p-value. State a conclusion in the context of the problem, using a significance level of $\alpha = 0.05$.

```
survdiff(Surv(Years, Censor) ~ Gender,
                  data = graduate)

## Call:
## survdiff(formula = Surv(Years, Censor) ~ Gender, data = graduate)
##
##             N Observed Expected (O-E)^2/E (O-E)^2/V
## Gender=0 472      273      302      2.78      6.11
## Gender=1 528      341      312      2.69      6.11
##
##  Chisq= 6.1  on 1 degrees of freedom, p= 0.0135
```

$\chi^2_1 = 6.1$ and the p-value=.0135. At the 0.06 level, we have evidence that the probability of not graduating does differ at some time t among males and females. We have evidence that time to graduate is associated with gender.

3. `Minitab` Locate the Minitab file `Melt Times Lab 5 W2018` on PolyLearn. We are interested in comparing the "survival" experiences of milk chocolate and white chocolate chips.

   (a) Briefly describe the two populations of interest relevant to this (pseudo) experiment.

   Population 1: All milk chocolate chips.

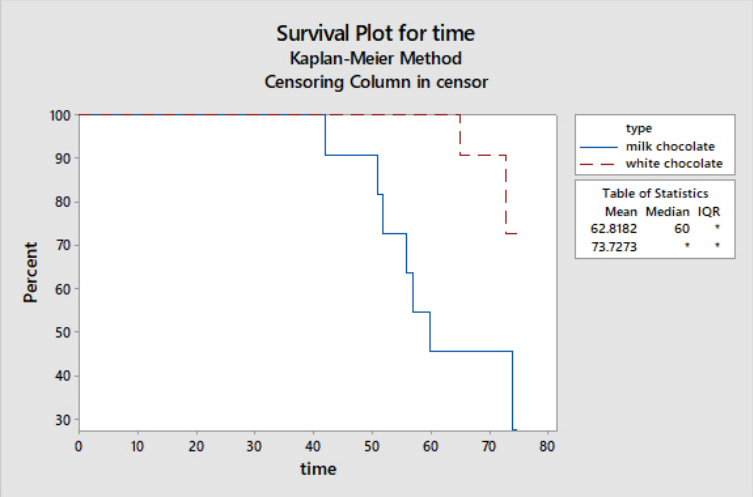
   Population 2: All white chocolate chips.

   (b) State the null and alternative hypotheses in words and using symbols to compare the survival experiences of the two appropriate populations described.

   $H_0 : S_1(t) = S_2(t)$ for all t. The probability of not melting after any time t does not depend on the type of chip, i.e. whether it is milk or white chocolate for all time t.

   $H_A : S_1(t) \neq S_2(t)$ for some t. The probability of not melting after any time t does depend on the type of chip, i.e. whether it is milk or dark chocolate for some time t.

   (c) Plot the Kaplan-Meier curves for both the milk chocolate and white chocolate melting times. In the context of the problem, compare the survival experiences of the two samples of chips. Does there appear to be any observed differences in the survival experiences?



milk chocolate chips tend to experience a complete melt quicker than white chocolate chips

   (d) Report the median event time (if available) for each type of chip to melt, and comment on which type of chip tends to melt faster.

milk chocolate: median = 60

white chocolate: median = not estimable, though can be assumed to be greater than 60

milk chocolate tends to melt faster

(e) Conduct the log-rank and Wilcoxon tests at the .05 significance level (the output for both tests follows the Kaplan-Meier results in the Session Window). Be sure to report the values of the test statistics and p-values, and state a conclusion in the context of the problem. If the results of the tests disagree, attempt to explain why.

```
Test Statistics
Method    Chi-Square  DF P-Value
Log-Rank  5.32688     1  0.021
Wilcoxon  5.73901     1  0.017
```

At the $\alpha = 0.05$ level of significance, we do have sufficient evidence to reject Ho. There is evidence that the probability of not melting after any time t depends on chip type for at least one time point. There is evidence of an association between chip "survival" and chip type.