K-M Life Table

Huanyu Chen

2023-12-08

```
dat <- read_csv("./data/heart_failure.csv") %>%
  arrange(TIME) %>% janitor::clean_names()
```

```
## Rows: 299 Columns: 13
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (13): TIME, Event, Gender, Smoking, Diabetes, BP, Anaemia, Age, Ejection...
##
## 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.
```

```
dat
```

```
## # A tibble: 299 x 13
##     time event gender smoking diabetes    bp anaemia   age ejection_fraction
##    <dbl> <dbl>  <dbl>   <dbl>    <dbl> <dbl>   <dbl> <dbl>             <dbl>
##  1     4     1      1       0        0     1       0    75                20
##  2     6     1      1       0        0     0       0    55                38
##  3     7     1      1       1        0     0       0    65                20
##  4     7     1      1       0        0     0       1    50                20
##  5     8     1      0       0        1     0       1    65                20
##  6     8     1      1       1        0     1       1    90                40
##  7    10     1      1       0        0     0       1    75                15
##  8    10     1      1       1        1     0       1    60                60
##  9    10     1      0       0        0     0       0    65                65
## 10    10     1      1       1        0     1       1    80                35
## # i 289 more rows
## # i 4 more variables: sodium <dbl>, creatinine <dbl>, pletelets <dbl>,
## #   cpk <dbl>
```

Fleming-Harrington Survival Estimate

$$\hat{S}_F(t) = \begin{cases} \prod_{t_i \le t} exp\left(-\frac{d_i}{n_i}\right) & \text{if } t > t_1 \\ 1 & \text{otherwise} \end{cases}$$

```
surv_object <- Surv(time = dat$time, event = dat$event)

# Calculate Fleming-Harrington table
```

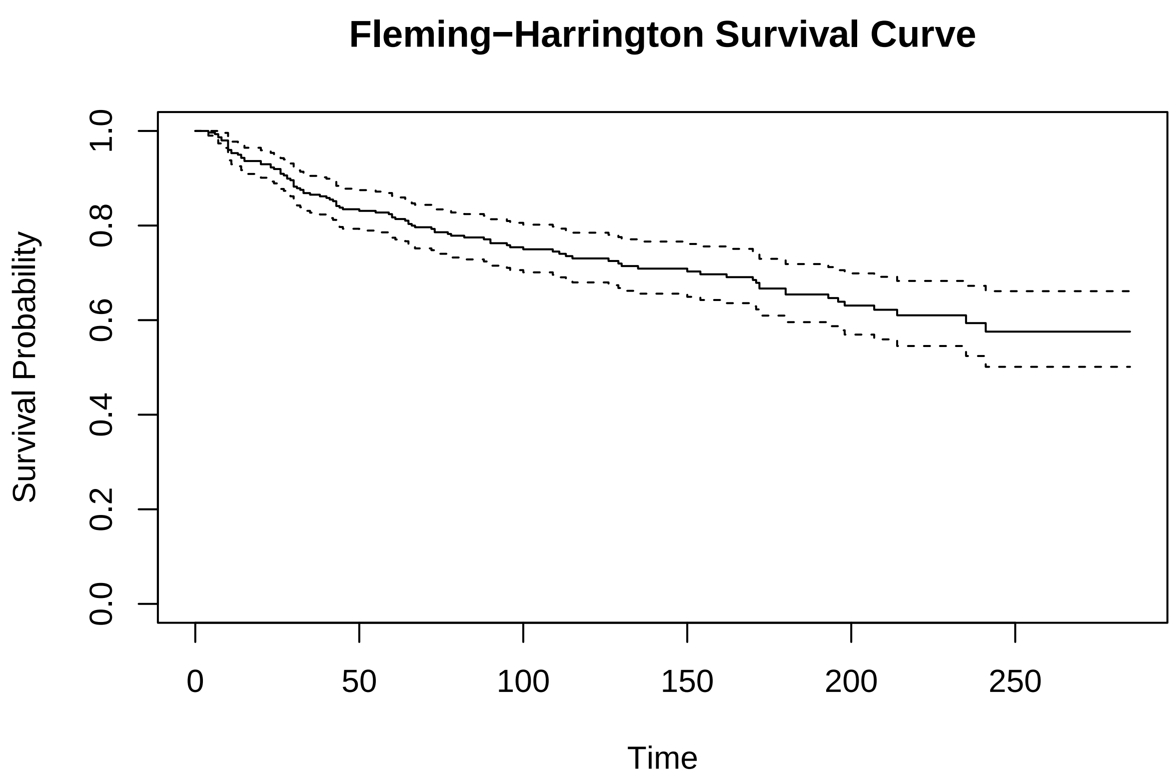
```r
fh_table <- survfit(surv_object ~ 1)

# Extract survival times and number of events
time_points <- fh_table$time
n_events <- fh_table$n.event
n_at_risk <- fh_table$n.risk

# Calculate Fleming-Harrington weights
fh_weights <- cumprod(exp(n_events / n_at_risk))

# Create a data frame for Fleming-Harrington statistics
fh_stats <- data.frame(Time = time_points,
                       N_events = n_events, N_at_risk = n_at_risk,
                       FH_Weights = fh_weights)

# Plotting the Fleming-Harrington survival curve
plot(fh_table, xlab = "Time", ylab = "Survival Probability",
     main = "Fleming-Harrington Survival Curve")
```



Kaplan-Meier Survival Estimate

$$\hat{S}_K(t) = \begin{cases} \prod_{t_i \le t} \left(1 - \frac{d_i}{n_i}\right) & \text{if } t > t_1 \\ 1 & \text{otherwise} \end{cases}$$

```r
surv_object <- Surv(time = dat$time, event = dat$event)
km_fit <- survfit(surv_object ~ 1)
summary(km_fit)
```

```
## Call: survfit(formula = surv_object ~ 1)
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##     4    299       1    0.997 0.00334        0.990        1.000
##     6    298       1    0.993 0.00471        0.984        1.000
##     7    297       2    0.987 0.00664        0.974        1.000
##     8    295       2    0.980 0.00811        0.964        0.996
##    10    293       6    0.960 0.01135        0.938        0.982
##    11    287       2    0.953 0.01222        0.930        0.977
##    13    284       1    0.950 0.01263        0.925        0.975
##    14    283       2    0.943 0.01340        0.917        0.970
##    15    281       2    0.936 0.01412        0.909        0.964
##    20    278       2    0.930 0.01480        0.901        0.959
##    23    275       2    0.923 0.01545        0.893        0.954
##    24    273       1    0.920 0.01575        0.889        0.951
##    26    272       3    0.909 0.01663        0.877        0.943
##    27    269       1    0.906 0.01691        0.873        0.940
##    28    268       2    0.899 0.01745        0.866        0.934
##    29    266       1    0.896 0.01771        0.862        0.931
##    30    264       4    0.882 0.01869        0.846        0.920
##    31    259       1    0.879 0.01893        0.843        0.917
##    32    258       1    0.875 0.01916        0.839        0.914
##    33    257       2    0.869 0.01961        0.831        0.908
##    35    254       1    0.865 0.01983        0.827        0.905
##    38    253       1    0.862 0.02004        0.823        0.902
##    40    252       1    0.858 0.02025        0.820        0.899
##    41    251       1    0.855 0.02046        0.816        0.896
##    42    250       1    0.852 0.02066        0.812        0.893
##    43    249       3    0.841 0.02124        0.801        0.884
##    44    246       1    0.838 0.02143        0.797        0.881
##    45    245       1    0.834 0.02161        0.793        0.878
##    50    244       1    0.831 0.02179        0.789        0.875
##    55    241       1    0.828 0.02197        0.786        0.872
##    59    240       1    0.824 0.02215        0.782        0.869
##    60    239       2    0.817 0.02250        0.774        0.863
##    61    236       1    0.814 0.02267        0.771        0.859
##    64    234       1    0.810 0.02283        0.767        0.856
##    65    233       2    0.803 0.02316        0.759        0.850
##    66    231       1    0.800 0.02332        0.755        0.847
##    67    230       1    0.796 0.02348        0.752        0.844
##    72    227       1    0.793 0.02364        0.748        0.841
##    73    225       2    0.786 0.02394        0.740        0.834
##    77    217       1    0.782 0.02411        0.736        0.831
##    78    216       1    0.779 0.02427        0.732        0.828
##    82    207       1    0.775 0.02444        0.728        0.824
##    88    194       1    0.771 0.02464        0.724        0.821
##    90    189       2    0.763 0.02504        0.715        0.813
##    95    180       1    0.758 0.02526        0.711        0.810
##    96    175       1    0.754 0.02548        0.706        0.806
##   100    173       1    0.750 0.02571        0.701        0.802
##   109    159       1    0.745 0.02597        0.696        0.798
##   111    155       1    0.740 0.02625        0.691        0.794
```

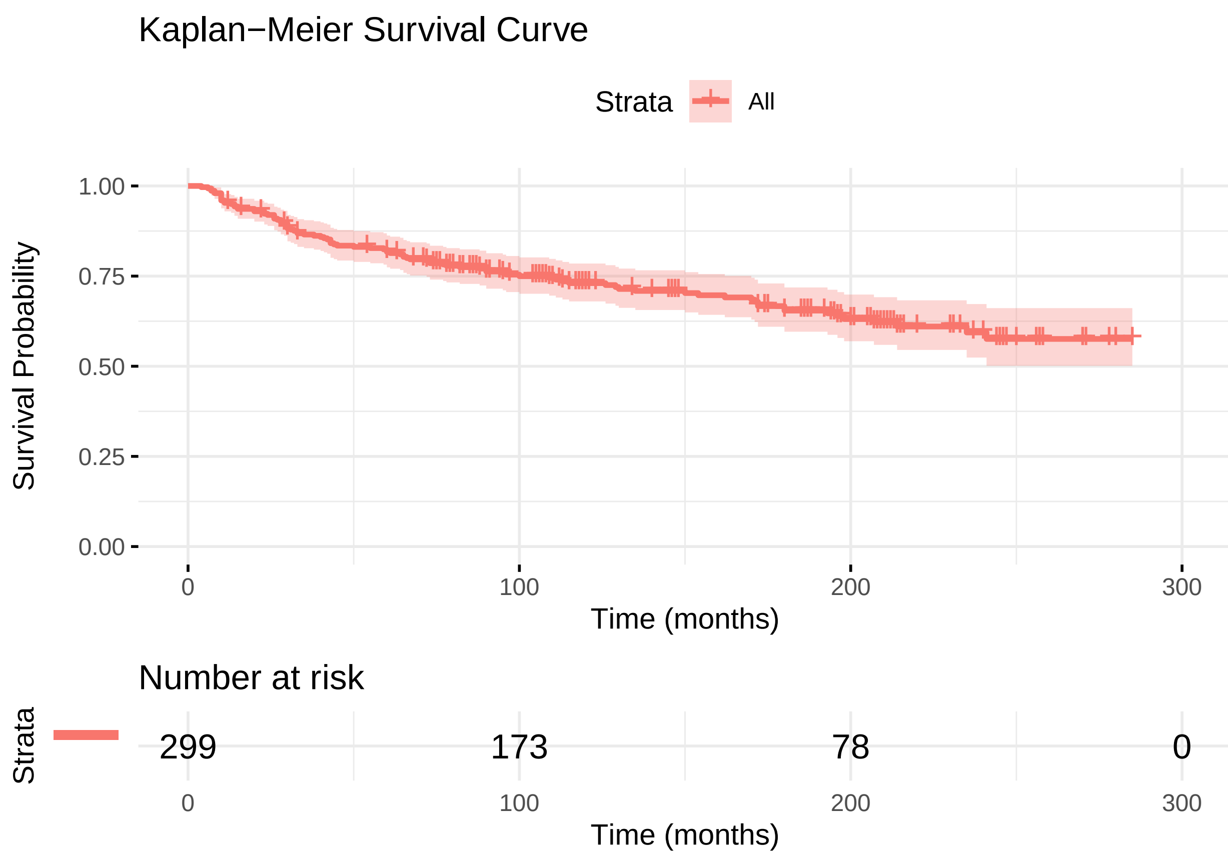
```
##    113    152       1    0.735 0.02652        0.685        0.789
##    115    150       1    0.730 0.02679        0.680        0.785
##    126    136       1    0.725 0.02713        0.674        0.780
##    129    135       1    0.720 0.02746        0.668        0.776
##    130    134       1    0.714 0.02777        0.662        0.771
##    135    132       1    0.709 0.02808        0.656        0.766
##    150    118       1    0.703 0.02848        0.649        0.761
##    154    117       1    0.697 0.02886        0.643        0.756
##    162    116       1    0.691 0.02923        0.636        0.751
##    170    115       1    0.685 0.02959        0.629        0.745
##    171    114       1    0.679 0.02993        0.623        0.740
##    172    113       2    0.667 0.03059        0.610        0.730
##    180    106       2    0.654 0.03128        0.596        0.719
##    193     86       1    0.647 0.03183        0.587        0.712
##    196     83       1    0.639 0.03238        0.578        0.706
##    198     79       1    0.631 0.03297        0.569        0.699
##    207     71       1    0.622 0.03368        0.559        0.692
##    214     53       1    0.610 0.03503        0.545        0.683
##    235     37       1    0.594 0.03776        0.524        0.673
##    241     33       1    0.576 0.04068        0.501        0.661
```

Life Table

```
surv_object <- Surv(time = dat$time, event = dat$event)
life_table <- survfit(surv_object ~ 1)
summary(life_table, times = c(1:10)*30) # Monthly intervals
```

```
## Call: survfit(formula = surv_object ~ 1)
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    30    264      35    0.882  0.0187        0.846        0.920
##    60    239      19    0.817  0.0225        0.774        0.863
##    90    189      15    0.763  0.0250        0.715        0.813
##   120    145       7    0.730  0.0268        0.680        0.785
##   150    118       5    0.703  0.0285        0.649        0.761
##   180    106       8    0.654  0.0313        0.596        0.719
##   210     62       4    0.622  0.0337        0.559        0.692
##   240     34       2    0.594  0.0378        0.524        0.673
##   270      6       1    0.576  0.0407        0.501        0.661
```

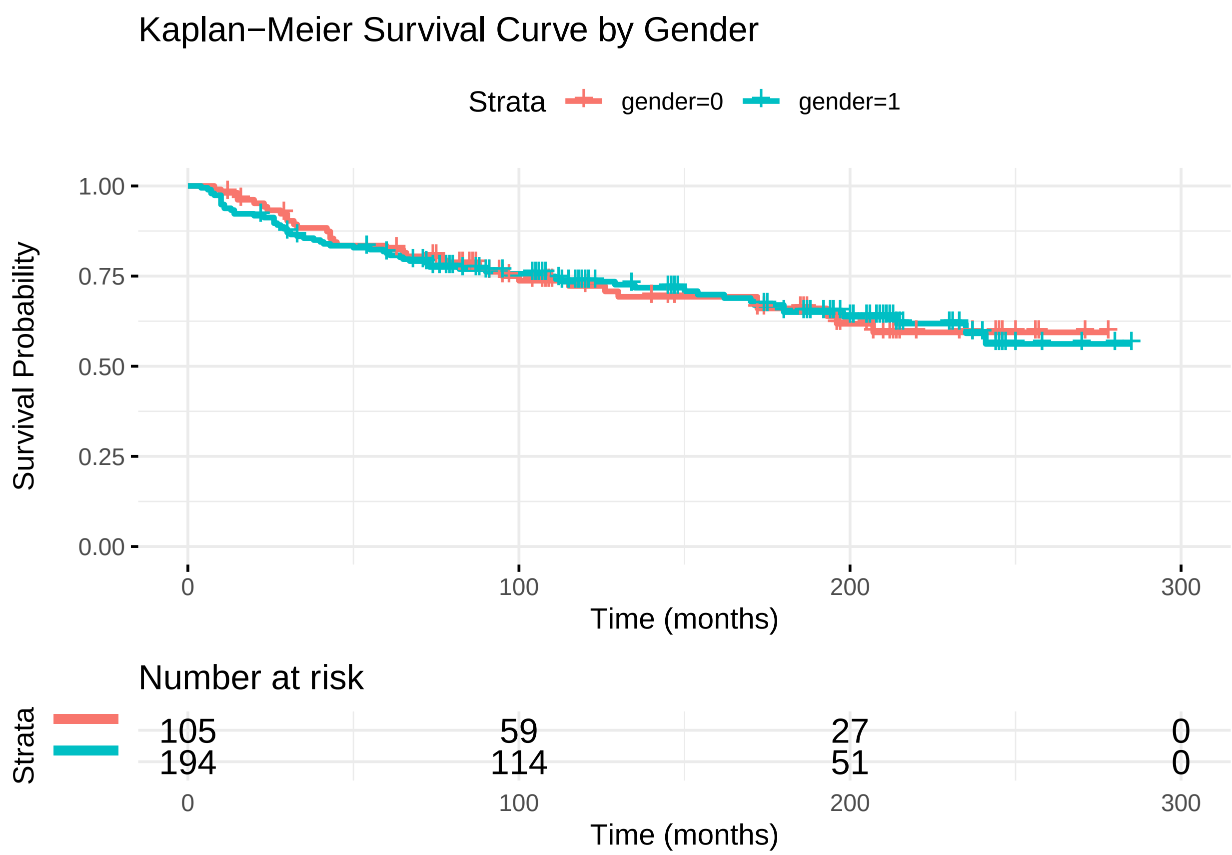
```
ggsurvplot(km_fit, data = dat, title = "Kaplan-Meier Survival Curve",
           xlab = "Time (months)", ylab = "Survival Probability",
           risk.table = TRUE, risk.table.y.text = FALSE,
           ggtheme = theme_minimal())
```



Life Table Stratified by Gender

Plot

```
surv_object <- Surv(time = dat$time, event = dat$event)
km_fit_gender <- survfit(surv_object ~ gender, data = dat)
ggsurvplot(km_fit_gender, data = dat,
           title = "Kaplan-Meier Survival Curve by Gender",
           xlab = "Time (months)", ylab = "Survival Probability",
           risk.table = TRUE, risk.table.y.text = FALSE,
           ggtheme = theme_minimal())
```



Log Rank Test

We would like to test if there is difference in survival between genders.

```
logrank_test <- survdiff(surv_object ~ gender, data = dat)
print(logrank_test)
```

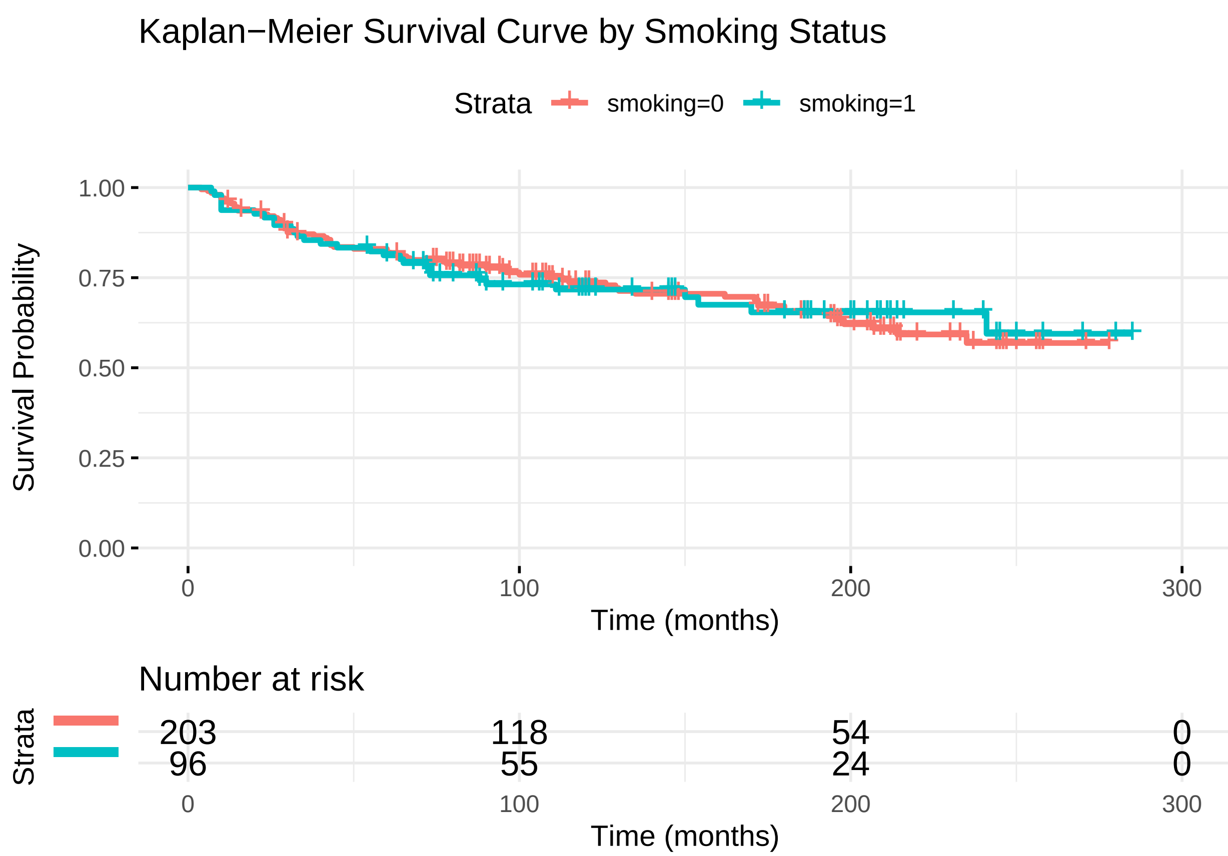
```
## Call:
## survdiff(formula = surv_object ~ gender, data = dat)
##
##             N Observed Expected (O-E)^2/E (O-E)^2/V
## gender=0 105       34     34.3   0.00254   0.00397
## gender=1 194       62     61.7   0.00141   0.00397
##
##  Chisq= 0  on 1 degrees of freedom, p= 0.9
```

With a p-value of 0.9, there is no sufficient evidence to reject the null hypothesis. In this case, there is no difference in survival between the two genders based on this dataset.

Life Table Stratified by Smoking Status

Plot

```
surv_object <- Surv(time = dat$time, event = dat$event)
km_fit_smoking <- survfit(surv_object ~ smoking, data = dat)
ggsurvplot(km_fit_smoking, data = dat,
           title = "Kaplan-Meier Survival Curve by Smoking Status",
           xlab = "Time (months)", ylab = "Survival Probability",
           risk.table = TRUE, risk.table.y.text = FALSE,
           ggtheme = theme_minimal())
```



Log Rank Test

We would like to test if there is difference in survival between smoking status.

```
logrank_test <- survdiff(surv_object ~ smoking, data = dat)
print(logrank_test)
```

```
## Call:
## survdiff(formula = surv_object ~ smoking, data = dat)
##
##               N Observed Expected (O-E)^2/E (O-E)^2/V
```

```
## smoking=0 203       66     65.8   0.00064   0.00204
## smoking=1  96       30     30.2   0.00139   0.00204
##
##  Chisq= 0  on 1 degrees of freedom, p= 1
```

With a p-value of 1, there is no sufficient evidence to reject the null hypothesis. In this case, there is no difference in survival between smokers and non-smokers based on this dataset.