GR5291 Advanced Data Analysis Problem Set Survival Analysis

Francis Zhang

November 22, 2024

Question

Consider the Mayo Clinic Lung Cancer Data in R package survival : data(lung) or data(cancer): including the variables

```
inst:       Institution code

time:       Survival time in days

status:     censoring status 1=censored, 2=dead

age:        Age in years

sex:        Male=1 Female=2, etc.
```

1.Estimate and plot the survival curve for the combined Male and Female data using the following methods:

a.Kaplan-Meier

b.Fleming-Harrington

2.Estimate the median survival time, using the estimated survival curves from 1a and 1b.

3.Using a log-rank test, compare the survival distributions for Male and Female

Solution

Question 1

```r
# Load necessary libraries
library(survival)
library(survminer)
```

```
## Loading required package: ggplot2

## Loading required package: ggpubr

##
## Attaching package: 'survminer'

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

```r
# Load the lung dataset
data(lung)
```

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

```r
head(lung)
```

```
##   inst time status age sex ph.ecog ph.karno pat.karno meal.cal wt.loss
## 1    3  306      2  74   1       1       90       100     1175      NA
## 2    3  455      2  68   1       0       90        90     1225      15
## 3    3 1010      1  56   1       0       90        90       NA      15
## 4    5  210      2  57   1       1       90        60     1150      11
## 5    1  883      2  60   1       0      100        90       NA       0
## 6   12 1022      1  74   1       1       50        80      513       0
```

```r
# Create a survival object
surv_object <- Surv(time = lung$time, event = lung$status == 2)

# Fit the Kaplan-Meier survival model
km_fit <- survfit(surv_object ~ 1, data = lung)
summary(km_fit)
```

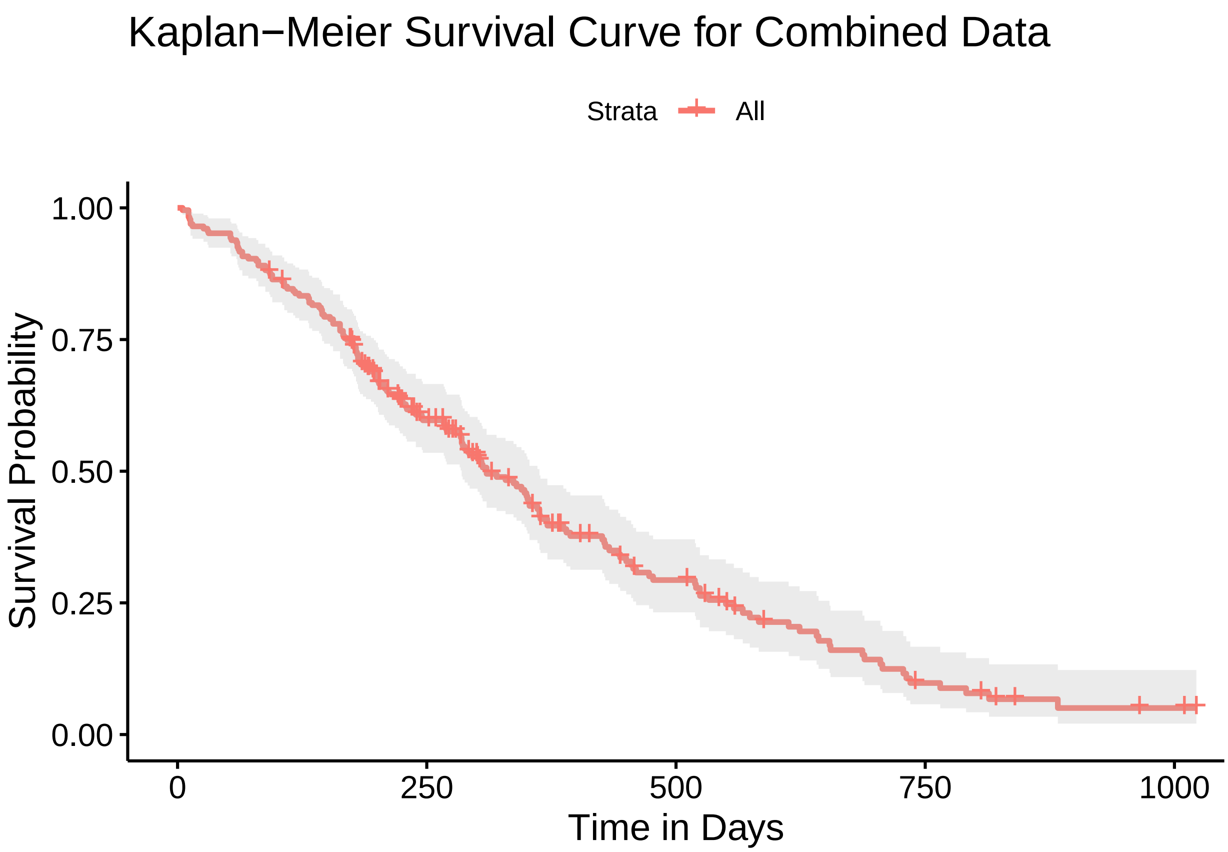
a.Kaplan-Meier

```
## Call: survfit(formula = surv_object ~ 1, data = lung)
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##     5    228       1   0.9956 0.00438       0.9871        1.000
##    11    227       3   0.9825 0.00869       0.9656        1.000
##    12    224       1   0.9781 0.00970       0.9592        0.997
##    13    223       2   0.9693 0.01142       0.9472        0.992
##    15    221       1   0.9649 0.01219       0.9413        0.989
##    26    220       1   0.9605 0.01290       0.9356        0.986
##    30    219       1   0.9561 0.01356       0.9299        0.983
##    31    218       1   0.9518 0.01419       0.9243        0.980
##    53    217       2   0.9430 0.01536       0.9134        0.974
##    54    215       1   0.9386 0.01590       0.9079        0.970
##    59    214       1   0.9342 0.01642       0.9026        0.967
##    60    213       2   0.9254 0.01740       0.8920        0.960
##    61    211       1   0.9211 0.01786       0.8867        0.957
##    62    210       1   0.9167 0.01830       0.8815        0.953
##    65    209       2   0.9079 0.01915       0.8711        0.946
##    71    207       1   0.9035 0.01955       0.8660        0.943
##    79    206       1   0.8991 0.01995       0.8609        0.939
##    81    205       2   0.8904 0.02069       0.8507        0.932
##    88    203       2   0.8816 0.02140       0.8406        0.925
##    92    201       1   0.8772 0.02174       0.8356        0.921
##    93    199       1   0.8728 0.02207       0.8306        0.917
##    95    198       2   0.8640 0.02271       0.8206        0.910
##   105    196       1   0.8596 0.02302       0.8156        0.906
##   107    194       2   0.8507 0.02362       0.8056        0.898
##   110    192       1   0.8463 0.02391       0.8007        0.894
##   116    191       1   0.8418 0.02419       0.7957        0.891
##   118    190       1   0.8374 0.02446       0.7908        0.887
##   122    189       1   0.8330 0.02473       0.7859        0.883
##   131    188       1   0.8285 0.02500       0.7810        0.879
##   132    187       2   0.8197 0.02550       0.7712        0.871
##   135    185       1   0.8153 0.02575       0.7663        0.867
```

```
##   142    184        1   0.8108 0.02598        0.7615        0.863
##   144    183        1   0.8064 0.02622        0.7566        0.859
##   145    182        2   0.7975 0.02667        0.7469        0.852
##   147    180        1   0.7931 0.02688        0.7421        0.848
##   153    179        1   0.7887 0.02710        0.7373        0.844
##   156    178        2   0.7798 0.02751        0.7277        0.836
##   163    176        3   0.7665 0.02809        0.7134        0.824
##   166    173        2   0.7577 0.02845        0.7039        0.816
##   167    171        1   0.7532 0.02863        0.6991        0.811
##   170    170        1   0.7488 0.02880        0.6944        0.807
##   175    167        1   0.7443 0.02898        0.6896        0.803
##   176    165        1   0.7398 0.02915        0.6848        0.799
##   177    164        1   0.7353 0.02932        0.6800        0.795
##   179    162        2   0.7262 0.02965        0.6704        0.787
##   180    160        1   0.7217 0.02981        0.6655        0.783
##   181    159        2   0.7126 0.03012        0.6559        0.774
##   182    157        1   0.7081 0.03027        0.6511        0.770
##   183    156        1   0.7035 0.03041        0.6464        0.766
##   186    154        1   0.6989 0.03056        0.6416        0.761
##   189    152        1   0.6943 0.03070        0.6367        0.757
##   194    149        1   0.6897 0.03085        0.6318        0.753
##   197    147        1   0.6850 0.03099        0.6269        0.749
##   199    145        1   0.6803 0.03113        0.6219        0.744
##   201    144        2   0.6708 0.03141        0.6120        0.735
##   202    142        1   0.6661 0.03154        0.6071        0.731
##   207    139        1   0.6613 0.03168        0.6020        0.726
##   208    138        1   0.6565 0.03181        0.5970        0.722
##   210    137        1   0.6517 0.03194        0.5920        0.717
##   212    135        1   0.6469 0.03206        0.5870        0.713
##   218    134        1   0.6421 0.03218        0.5820        0.708
##   222    132        1   0.6372 0.03231        0.5769        0.704
##   223    130        1   0.6323 0.03243        0.5718        0.699
##   226    126        1   0.6273 0.03256        0.5666        0.694
##   229    125        1   0.6223 0.03268        0.5614        0.690
##   230    124        1   0.6172 0.03280        0.5562        0.685
##   239    121        2   0.6070 0.03304        0.5456        0.675
##   245    117        1   0.6019 0.03316        0.5402        0.670
##   246    116        1   0.5967 0.03328        0.5349        0.666
##   267    112        1   0.5913 0.03341        0.5294        0.661
##   268    111        1   0.5860 0.03353        0.5239        0.656
##   269    110        1   0.5807 0.03364        0.5184        0.651
##   270    108        1   0.5753 0.03376        0.5128        0.645
##   283    104        1   0.5698 0.03388        0.5071        0.640
##   284    103        1   0.5642 0.03400        0.5014        0.635
##   285    101        2   0.5531 0.03424        0.4899        0.624
##   286     99        1   0.5475 0.03434        0.4841        0.619
##   288     98        1   0.5419 0.03444        0.4784        0.614
##   291     97        1   0.5363 0.03454        0.4727        0.608
##   293     94        1   0.5306 0.03464        0.4669        0.603
##   301     91        1   0.5248 0.03475        0.4609        0.597
##   303     89        1   0.5189 0.03485        0.4549        0.592
##   305     87        1   0.5129 0.03496        0.4488        0.586
##   306     86        1   0.5070 0.03506        0.4427        0.581
##   310     85        2   0.4950 0.03523        0.4306        0.569
```

```
##   320     82        1   0.4890 0.03532       0.4244         0.563
##   329     81        1   0.4830 0.03539       0.4183         0.558
##   337     79        1   0.4768 0.03547       0.4121         0.552
##   340     78        1   0.4707 0.03554       0.4060         0.546
##   345     77        1   0.4646 0.03560       0.3998         0.540
##   348     76        1   0.4585 0.03565       0.3937         0.534
##   350     75        1   0.4524 0.03569       0.3876         0.528
##   351     74        1   0.4463 0.03573       0.3815         0.522
##   353     73        2   0.4340 0.03578       0.3693         0.510
##   361     70        1   0.4278 0.03581       0.3631         0.504
##   363     69        2   0.4154 0.03583       0.3508         0.492
##   364     67        1   0.4092 0.03582       0.3447         0.486
##   371     65        2   0.3966 0.03581       0.3323         0.473
##   387     60        1   0.3900 0.03582       0.3258         0.467
##   390     59        1   0.3834 0.03582       0.3193         0.460
##   394     58        1   0.3768 0.03580       0.3128         0.454
##   426     55        1   0.3700 0.03580       0.3060         0.447
##   428     54        1   0.3631 0.03579       0.2993         0.440
##   429     53        1   0.3563 0.03576       0.2926         0.434
##   433     52        1   0.3494 0.03573       0.2860         0.427
##   442     51        1   0.3426 0.03568       0.2793         0.420
##   444     50        1   0.3357 0.03561       0.2727         0.413
##   450     48        1   0.3287 0.03555       0.2659         0.406
##   455     47        1   0.3217 0.03548       0.2592         0.399
##   457     46        1   0.3147 0.03539       0.2525         0.392
##   460     44        1   0.3076 0.03530       0.2456         0.385
##   473     43        1   0.3004 0.03520       0.2388         0.378
##   477     42        1   0.2933 0.03508       0.2320         0.371
##   519     39        1   0.2857 0.03498       0.2248         0.363
##   520     38        1   0.2782 0.03485       0.2177         0.356
##   524     37        2   0.2632 0.03455       0.2035         0.340
##   533     34        1   0.2554 0.03439       0.1962         0.333
##   550     32        1   0.2475 0.03423       0.1887         0.325
##   558     30        1   0.2392 0.03407       0.1810         0.316
##   567     28        1   0.2307 0.03391       0.1729         0.308
##   574     27        1   0.2221 0.03371       0.1650         0.299
##   583     26        1   0.2136 0.03348       0.1571         0.290
##   613     24        1   0.2047 0.03325       0.1489         0.281
##   624     23        1   0.1958 0.03297       0.1407         0.272
##   641     22        1   0.1869 0.03265       0.1327         0.263
##   643     21        1   0.1780 0.03229       0.1247         0.254
##   654     20        1   0.1691 0.03188       0.1169         0.245
##   655     19        1   0.1602 0.03142       0.1091         0.235
##   687     18        1   0.1513 0.03090       0.1014         0.226
##   689     17        1   0.1424 0.03034       0.0938         0.216
##   705     16        1   0.1335 0.02972       0.0863         0.207
##   707     15        1   0.1246 0.02904       0.0789         0.197
##   728     14        1   0.1157 0.02830       0.0716         0.187
##   731     13        1   0.1068 0.02749       0.0645         0.177
##   735     12        1   0.0979 0.02660       0.0575         0.167
##   765     10        1   0.0881 0.02568       0.0498         0.156
##   791      9        1   0.0783 0.02462       0.0423         0.145
##   814      7        1   0.0671 0.02351       0.0338         0.133
##   883      4        1   0.0503 0.02285       0.0207         0.123
```

```r
# Plot the Kaplan-Meier survival curve
ggsurvkm <- ggsurvplot(km_fit, conf.int = TRUE,
                       title = "Kaplan-Meier Survival Curve for Combined Data",
                       xlab = "Time in Days", ylab = "Survival Probability")
ggsurvkm
```




```r
# Fit the Fleming-Harrington survival model
fh_fit <- survfit(surv_object ~ 1, type = "fh", data = lung)
summary(fh_fit)
```

b.Fleming-Harrington

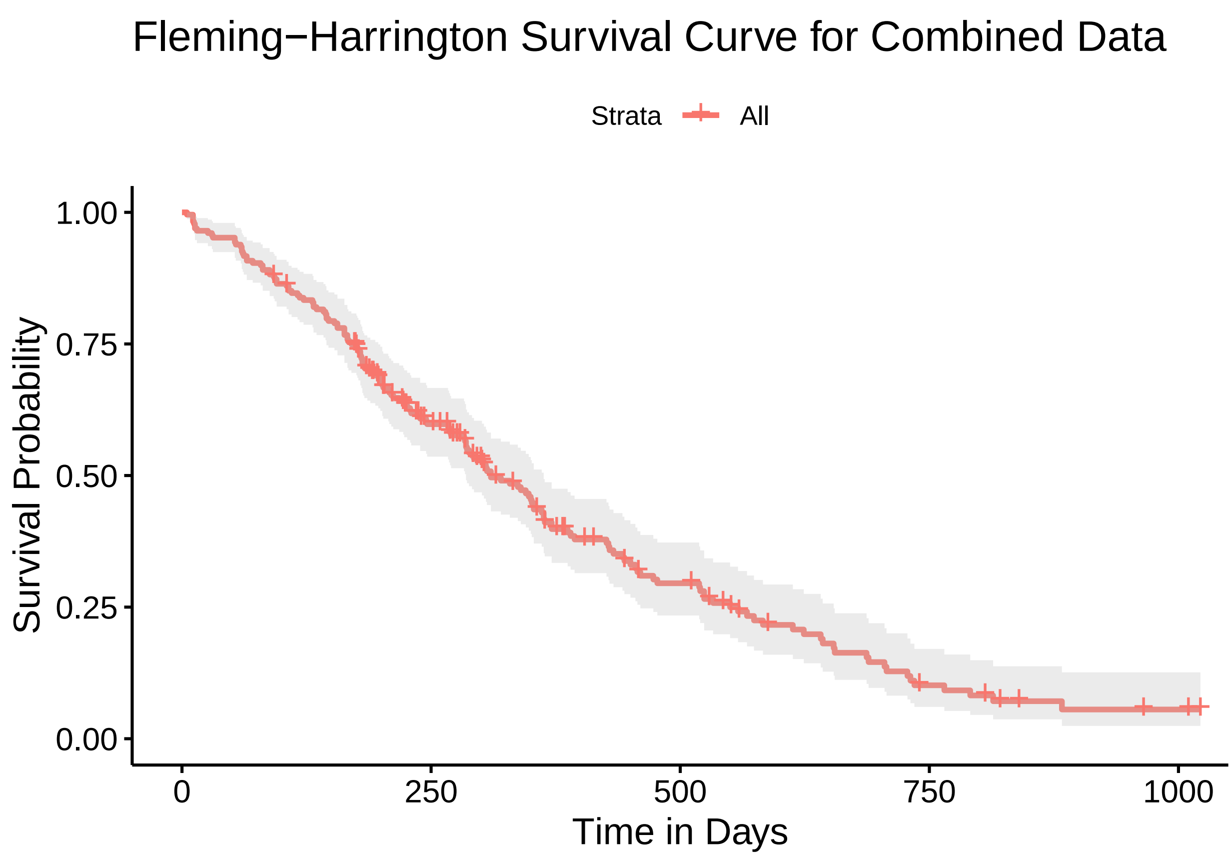
```
## Call: survfit(formula = surv_object ~ 1, data = lung, type = "fh")
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##     5    228       1   0.9956 0.00437       0.9871        1.000
##    11    227       3   0.9825 0.00868       0.9656        1.000
##    12    224       1   0.9781 0.00968       0.9593        0.997
##    13    223       2   0.9694 0.01140       0.9473        0.992
##    15    221       1   0.9650 0.01216       0.9414        0.989
##    26    220       1   0.9606 0.01287       0.9357        0.986
##    30    219       1   0.9562 0.01353       0.9301        0.983
##    31    218       1   0.9519 0.01416       0.9245        0.980
##    53    217       2   0.9431 0.01532       0.9135        0.974
##    54    215       1   0.9387 0.01587       0.9081        0.970
##    59    214       1   0.9344 0.01638       0.9028        0.967
```

```
##    60     213         2   0.9256 0.01736        0.8922         0.960
##    61     211         1   0.9212 0.01782        0.8870         0.957
##    62     210         1   0.9168 0.01827        0.8817         0.953
##    65     209         2   0.9081 0.01911        0.8714         0.946
##    71     207         1   0.9037 0.01951        0.8663         0.943
##    79     206         1   0.8993 0.01990        0.8612         0.939
##    81     205         2   0.8906 0.02065        0.8510         0.932
##    88     203         2   0.8818 0.02135        0.8410         0.925
##    92     201         1   0.8775 0.02169        0.8360         0.921
##    93     199         1   0.8731 0.02202        0.8309         0.917
##    95     198         2   0.8643 0.02266        0.8210         0.910
##   105     196         1   0.8599 0.02297        0.8160         0.906
##   107     194         2   0.8510 0.02357        0.8061         0.899
##   110     192         1   0.8466 0.02386        0.8011         0.895
##   116     191         1   0.8422 0.02414        0.7962         0.891
##   118     190         1   0.8378 0.02442        0.7913         0.887
##   122     189         1   0.8333 0.02468        0.7863         0.883
##   131     188         1   0.8289 0.02495        0.7814         0.879
##   132     187         2   0.8201 0.02545        0.7717         0.872
##   135     185         1   0.8157 0.02569        0.7668         0.868
##   142     184         1   0.8112 0.02593        0.7620         0.864
##   144     183         1   0.8068 0.02617        0.7571         0.860
##   145     182         2   0.7980 0.02662        0.7475         0.852
##   147     180         1   0.7936 0.02683        0.7427         0.848
##   153     179         1   0.7891 0.02705        0.7379         0.844
##   156     178         2   0.7803 0.02746        0.7283         0.836
##   163     176         3   0.7670 0.02804        0.7140         0.824
##   166     173         2   0.7582 0.02840        0.7045         0.816
##   167     171         1   0.7538 0.02858        0.6998         0.812
##   170     170         1   0.7493 0.02875        0.6951         0.808
##   175     167         1   0.7449 0.02892        0.6903         0.804
##   176     165         1   0.7404 0.02910        0.6855         0.800
##   177     164         1   0.7359 0.02927        0.6807         0.796
##   179     162         2   0.7268 0.02960        0.6711         0.787
##   180     160         1   0.7223 0.02976        0.6663         0.783
##   181     159         2   0.7132 0.03007        0.6567         0.775
##   182     157         1   0.7087 0.03022        0.6519         0.770
##   183     156         1   0.7042 0.03036        0.6471         0.766
##   186     154         1   0.6996 0.03050        0.6423         0.762
##   189     152         1   0.6950 0.03065        0.6375         0.758
##   194     149         1   0.6904 0.03079        0.6326         0.753
##   197     147         1   0.6857 0.03094        0.6277         0.749
##   199     145         1   0.6810 0.03108        0.6227         0.745
##   201     144         2   0.6716 0.03136        0.6128         0.736
##   202     142         1   0.6668 0.03149        0.6079         0.732
##   207     139         1   0.6621 0.03162        0.6029         0.727
##   208     138         1   0.6573 0.03176        0.5979         0.723
##   210     137         1   0.6525 0.03188        0.5929         0.718
##   212     135         1   0.6477 0.03201        0.5879         0.714
##   218     134         1   0.6429 0.03213        0.5829         0.709
##   222     132         1   0.6380 0.03225        0.5778         0.704
##   223     130         1   0.6331 0.03237        0.5728         0.700
##   226     126         1   0.6281 0.03250        0.5675         0.695
##   229     125         1   0.6231 0.03263        0.5623         0.690
```

```
##   230    124       1   0.6181 0.03275       0.5572        0.686
##   239    121       2   0.6079 0.03299       0.5466        0.676
##   245    117       1   0.6028 0.03311       0.5412        0.671
##   246    116       1   0.5976 0.03323       0.5359        0.666
##   267    112       1   0.5923 0.03335       0.5304        0.661
##   268    111       1   0.5870 0.03348       0.5249        0.656
##   269    110       1   0.5817 0.03359       0.5194        0.651
##   270    108       1   0.5763 0.03371       0.5139        0.646
##   283    104       1   0.5708 0.03383       0.5082        0.641
##   284    103       1   0.5653 0.03395       0.5025        0.636
##   285    101       2   0.5541 0.03418       0.4910        0.625
##   286     99       1   0.5486 0.03429       0.4853        0.620
##   288     98       1   0.5430 0.03439       0.4796        0.615
##   291     97       1   0.5374 0.03449       0.4739        0.609
##   293     94       1   0.5317 0.03459       0.4681        0.604
##   301     91       1   0.5259 0.03470       0.4621        0.599
##   303     89       1   0.5200 0.03480       0.4561        0.593
##   305     87       1   0.5141 0.03491       0.4500        0.587
##   306     86       1   0.5082 0.03501       0.4440        0.582
##   310     85       2   0.4963 0.03518       0.4319        0.570
##   320     82       1   0.4903 0.03527       0.4258        0.564
##   329     81       1   0.4842 0.03534       0.4197        0.559
##   337     79       1   0.4782 0.03542       0.4135        0.553
##   340     78       1   0.4721 0.03549       0.4074        0.547
##   345     77       1   0.4660 0.03555       0.4013        0.541
##   348     76       1   0.4599 0.03560       0.3951        0.535
##   350     75       1   0.4538 0.03565       0.3890        0.529
##   351     74       1   0.4477 0.03569       0.3829        0.523
##   353     73       2   0.4355 0.03574       0.3708        0.512
##   361     70       1   0.4293 0.03576       0.3647        0.505
##   363     69       2   0.4170 0.03578       0.3524        0.493
##   364     67       1   0.4108 0.03578       0.3463        0.487
##   371     65       2   0.3983 0.03577       0.3340        0.475
##   387     60       1   0.3917 0.03578       0.3275        0.468
##   390     59       1   0.3851 0.03578       0.3210        0.462
##   394     58       1   0.3785 0.03577       0.3145        0.456
##   426     55       1   0.3717 0.03577       0.3078        0.449
##   428     54       1   0.3649 0.03576       0.3011        0.442
##   429     53       1   0.3581 0.03573       0.2944        0.435
##   433     52       1   0.3512 0.03570       0.2878        0.429
##   442     51       1   0.3444 0.03565       0.2812        0.422
##   444     50       1   0.3376 0.03559       0.2746        0.415
##   450     48       1   0.3306 0.03553       0.2678        0.408
##   455     47       1   0.3237 0.03546       0.2611        0.401
##   457     46       1   0.3167 0.03537       0.2544        0.394
##   460     44       1   0.3096 0.03529       0.2476        0.387
##   473     43       1   0.3025 0.03519       0.2408        0.380
##   477     42       1   0.2954 0.03507       0.2340        0.373
##   519     39       1   0.2879 0.03497       0.2269        0.365
##   520     38       1   0.2804 0.03485       0.2198        0.358
##   524     37       2   0.2655 0.03456       0.2057        0.343
##   533     34       1   0.2578 0.03440       0.1984        0.335
##   550     32       1   0.2498 0.03425       0.1910        0.327
##   558     30       1   0.2416 0.03409       0.1833        0.319
```

```
## 567 28 1 0.2332 0.03393 0.1753 0.310
## 574 27 1 0.2247 0.03374 0.1674 0.302
## 583 26 1 0.2162 0.03352 0.1596 0.293
## 613 24 1 0.2074 0.03329 0.1514 0.284
## 624 23 1 0.1986 0.03302 0.1433 0.275
## 641 22 1 0.1897 0.03271 0.1353 0.266
## 643 21 1 0.1809 0.03236 0.1274 0.257
## 654 20 1 0.1721 0.03196 0.1196 0.248
## 655 19 1 0.1633 0.03152 0.1118 0.238
## 687 18 1 0.1544 0.03102 0.1042 0.229
## 689 17 1 0.1456 0.03048 0.0966 0.219
## 705 16 1 0.1368 0.02988 0.0892 0.210
## 707 15 1 0.1280 0.02923 0.0818 0.200
## 728 14 1 0.1192 0.02851 0.0745 0.190
## 731 13 1 0.1103 0.02773 0.0674 0.181
## 735 12 1 0.1015 0.02688 0.0604 0.171
## 765 10 1 0.0919 0.02600 0.0527 0.160
## 791 9 1 0.0822 0.02499 0.0453 0.149
## 814 7 1 0.0712 0.02394 0.0369 0.138
## 883 4 1 0.0555 0.02324 0.0244 0.126
```

```r
# Plot the Fleming-Harrington survival curve
ggsurvfh <- ggsurvplot(fh_fit, conf.int = TRUE, title =
                         "Fleming-Harrington Survival Curve for Combined Data",
                       xlab = "Time in Days", ylab = "Survival Probability")
ggsurvfh
```



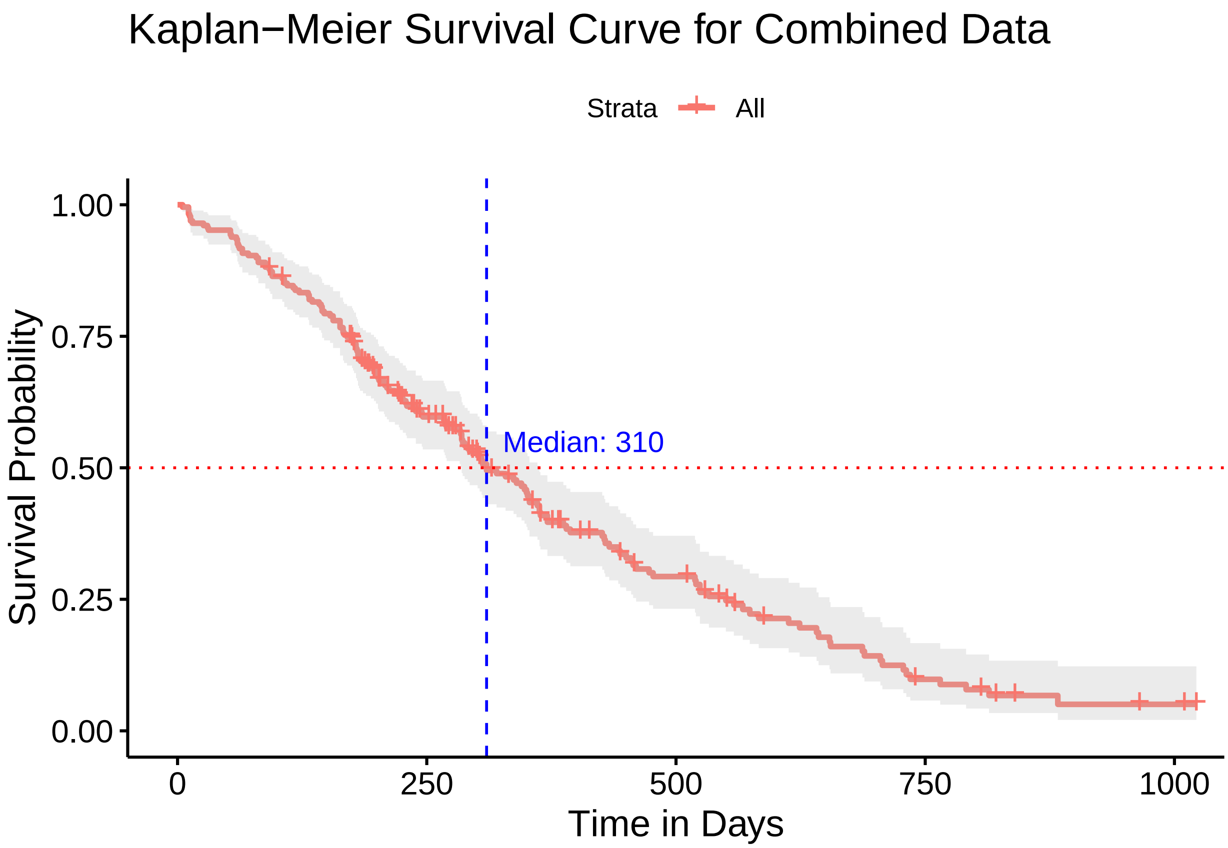
Question 2

Median survival means half of the people survived at that time. According to the summary tables above, median points should be around 310.

```
# Kaplan-Meier Median Survival Time:
km_median <- summary(km_fit)$table["median"]
km_median
```

```
## median
##    310
```

```
ggsurvkm$plot +
  geom_vline(xintercept = km_median, linetype = "dashed", color = "blue") +
  geom_hline(yintercept = 0.5, linetype = "dotted", color = "red") +
  annotate("text", x = km_median, y = 0.55, label =
             paste("Median:", round(km_median, 1)), color = "blue",
           hjust = -0.1)
```

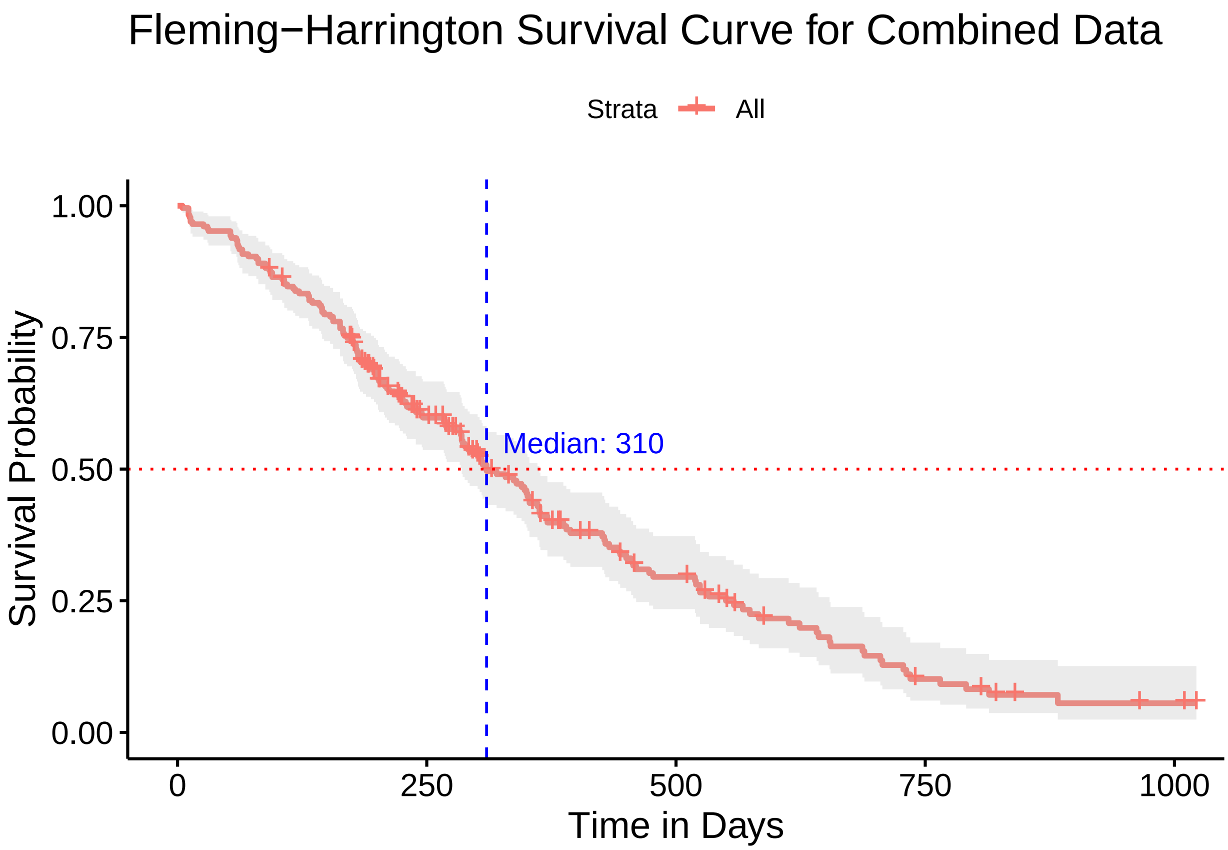



```
# Fleming-Harrington Median Survival Time:
fh_median <- summary(fh_fit)$table["median"]
fh_median
```

```
## median
##    310
```

```
ggsurvfh$plot +
  geom_vline(xintercept = fh_median, linetype = "dashed", color = "blue") +
  geom_hline(yintercept = 0.5, linetype = "dotted", color = "red") +
  annotate("text", x = fh_median, y = 0.55, label =
```

```r
     paste("Median:", round(fh_median, 1)), color = "blue",
    hjust = -0.1)
```



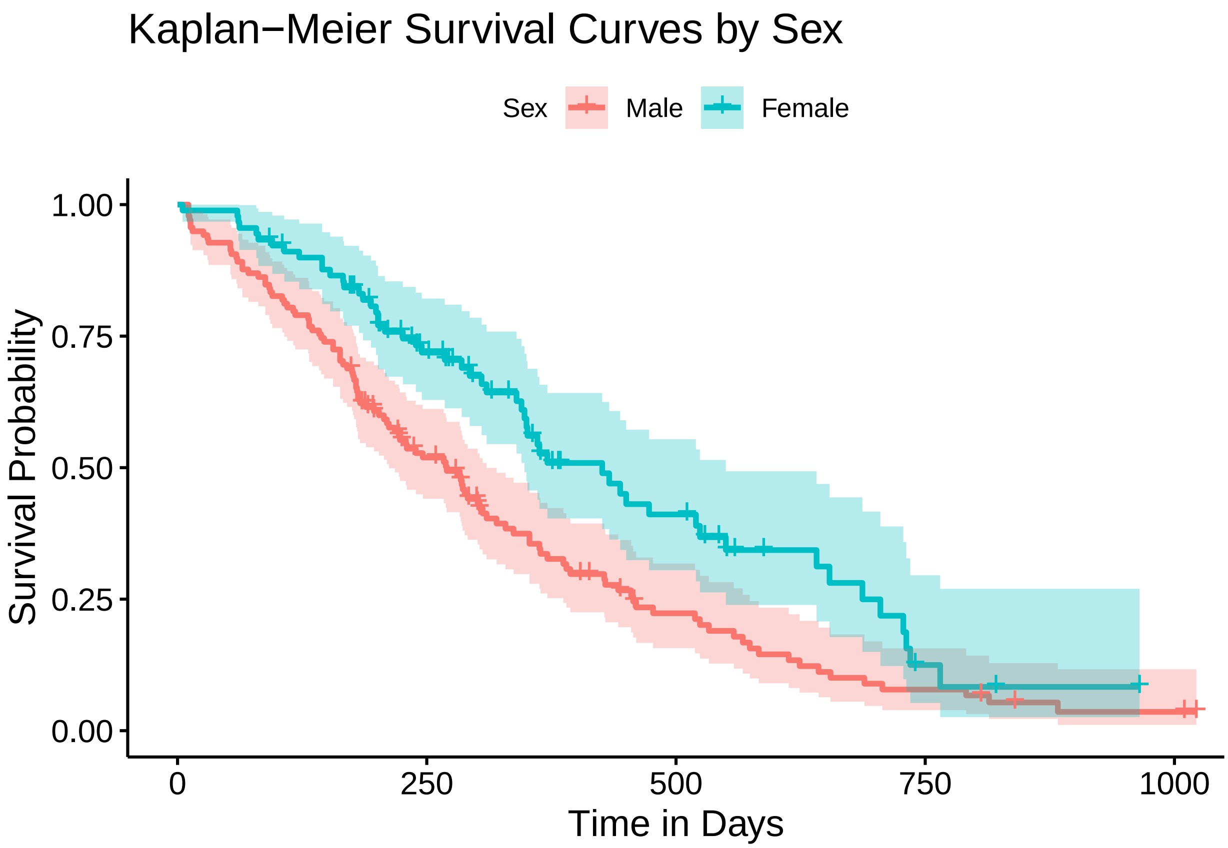
Question 3

The log-rank test is used to compare the survival distributions of two samples.

```r
# Convert sex to a factor variable
lung$sex <- factor(lung$sex, levels = c(1, 2), labels = c("Male", "Female"))

# Fit the Kaplan-Meier survival model by sex
km_sex_fit <- survfit(surv_object ~ sex, data = lung)

# Plot the Kaplan-Meier survival curves by sex
ggsurvplot(km_sex_fit, conf.int = TRUE,
           title = "Kaplan-Meier Survival Curves by Sex",
           xlab = "Time in Days", ylab = "Survival Probability",
           legend.title = "Sex", legend.labs = c("Male", "Female"))
```




```r
# Perform the log-rank test
log_rank_test <- survdiff(surv_object ~ sex, data = lung)
log_rank_test
```

```
## Call:
## survdiff(formula = surv_object ~ sex, data = lung)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## sex=Male   138      112     91.6      4.55      10.3
## sex=Female  90       53     73.4      5.68      10.3
##
##  Chisq= 10.3  on 1 degrees of freedom, p= 0.001
```

The Kaplan-Meier survival analysis indicates that females tend to have a higher survival probability compared to males over time. The log-rank test confirms a statistically significant difference in survival distributions between males and females ($\chi^2 = 10.3$, $p = 0.001$), suggesting that sex significantly impacts survival in this dataset.