Survival Analysis on Heart Failure

Lisa Nguyen

2024-05-09

Introduction

Heart failure is a medical condition characterized by the heart's inability to pump sufficient blood to meet the body's needs. This condition arises when the heart becomes too weak to effectively fill and distribute blood throughout the body. Heart failure is a serious condition that has no cure and requires proper medical treatment. According to the Centers for Disease Control and Prevention (2024), more than 6 million adults in the United States are affected by heart failure. Given the prevalence of heart failure as a chronic condition, the objective of this research project is to investigate which clinical features (i.e. health conditions and diagnoses) are evident in heart failure and how the survival of patients with heart failure change over time.

We will explore a dataset containing health failure clinical records, collected by the UC Irvine Machine Learning Repository. The dataset contains medical records of 299 patients who had heart failure during their follow-up period. Each patients' medical records contains 13 clinical features listed below (UCI Machine Learning Repository, 2020):

1) **age**: age of the patient (years)
2) **anaemia**: decrease of red blood cells or hemoglobin (boolean)
3) **creatinine phosphokinase (CPK)**: level of the CPK enzyme in the blood (mcg/L)
4) **diabetes**: if the patient has diabetes (boolean)
5) **ejection fraction**: percentage of blood leaving the heart at each contraction (percentage)
6) **high blood pressure**: if the patient has hypertension (boolean)
7) **platelets**: platelets in the blood (kiloplatelets/mL)
8) **sex**: woman or man (binary)
9) **serum creatinine**: level of serum creatinine in the blood (mg/dL)
10) **serum sodium**: level of serum sodium in the blood (mEq/L)
11) **smoking**: if the patient smokes or not (boolean)
12) **time**: follow-up period (days)
13) **[target] death event**: if the patient died during the follow-up period (boolean)

NOTE: The Survival Analysis is referenced from *Batra et al., (2023). *27 Survival Analysis*. The Epidemiologist R Handbook. https://epirhandbook.com/en/survival-analysis.html*.

Read and Load Data

To run and conduct survival analyses in R, it is important to install and load the `survival` package. This package provides functions such as the Kaplan-Meier estimation and Cox proportional hazards models.

We will also install and load the `survminer` package as it provides functions for visualizing survival analysis results, including the Kaplan-Meier curves, cumulative incidence curves, and forest plots for Cox models. We will discuss these functions later on.

Exploratory Data Analysis

Let's first explore our subjects in the study.

Table 1: Distribution of Age

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
40	51	60	60.83	70	95

The average age for patients that had heart failure is approximately 60 years old. The age distribution is slightly right skewed, indicating that most patients were in their late 40s to early 70s, or middle-aged.

Table 2: Count by Gender

Sex	Gender
Female	105
Male	194

There were more men that had heart failure compared to women. About 2/3 of patients were men.

Indicators of Heart Failure

Let's also explore possible indicators of heart failure based on our subjects.

Table 3: Distribution of Creatinine Phosphokinase (in mcg/L)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
23	116.5	250	581.8	582	7861

The normal range of CPK levels from 10 mcg/L to 120 mcg/L (Mount Sinai Health Clinic, 2024). Not only is the average level of CPK is 581.8 mcg/L, which is abnormally high, over half of the patients had CPK levels beyond the normal range. From this, high CPK levels can possibly be associated with heart failure.

Table 4: Distribution of Ejection Fraction (in %)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
14	30	38	38.08	45	80

A healthy range of ejection fraction is between 50% and 70% (Mayo Clinic, 2024). From our distribution, over half of the patients had lower ejection fraction, highlighting the heart's poor ability to pump blood. As a result, the ejection fraction can be seen as evidence of heart failure.

Table 5: Counts of Anaemia, Diabetes, and High Blood Pressure

Anaemia	Diabetes	High.Blood.Pressure
129	125	105

Approximately a third of patients had anaemia, diabetes, and/or hypertenion (i.e. high blood pressure), however, based on this there is not enough information to say these health conditions are linked to health failure.

Data Science Method: Survival Analysis

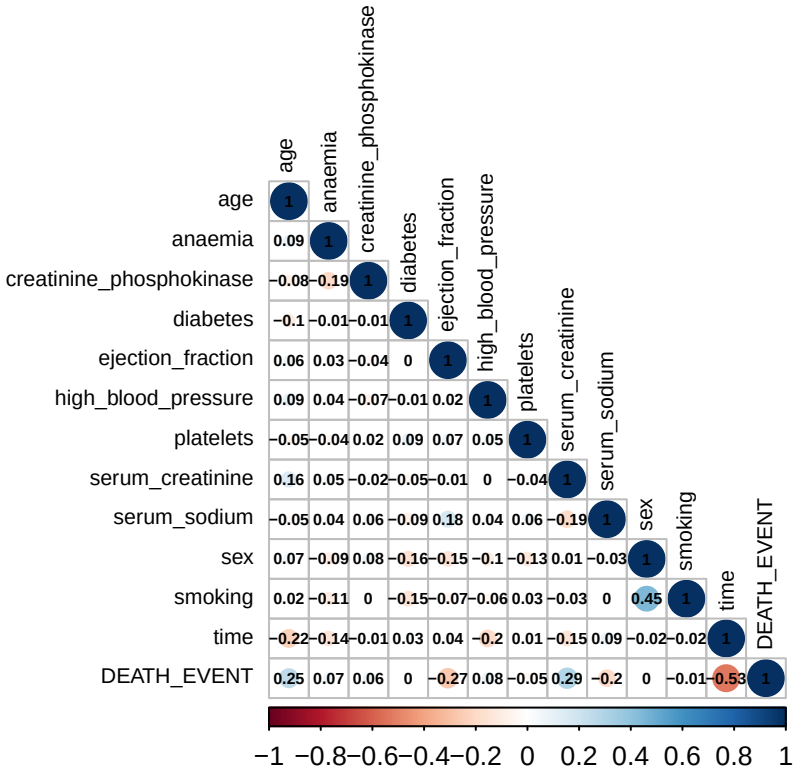
Survival analysis is a branch of statistics that describes time between an origin and an event. In other words, it analyzes the expected time of duration until an event of interest occurs. We often express this through survival probability, the probability that an event of interest has not occurred by a certain duration of time. Although survival analysis is used in several industries, it is widely used in medical research, so this is will be our focus.

We are essentially looking for 2 variables: (1) event variable (2) time variable

The event variable is known as the "failure", which deals with the occurrence of a disease, cure from a disease, death, etc. In this case, our event variable is whether the patient has died due to heart failure during the follow-up period. This is denoted by the "DEATH_EVENT" variable. Our time variable is known as the period of time that occurs before the patient died, or the follow-up time. Censoring also occurs at the end of the follow-up period when the individuals have not had the event of interest; hence, the time to event is unknown and the DEATH_EVENT will equal 0. This means we have some information about the individual survival time, but we don't know the survival time exactly.

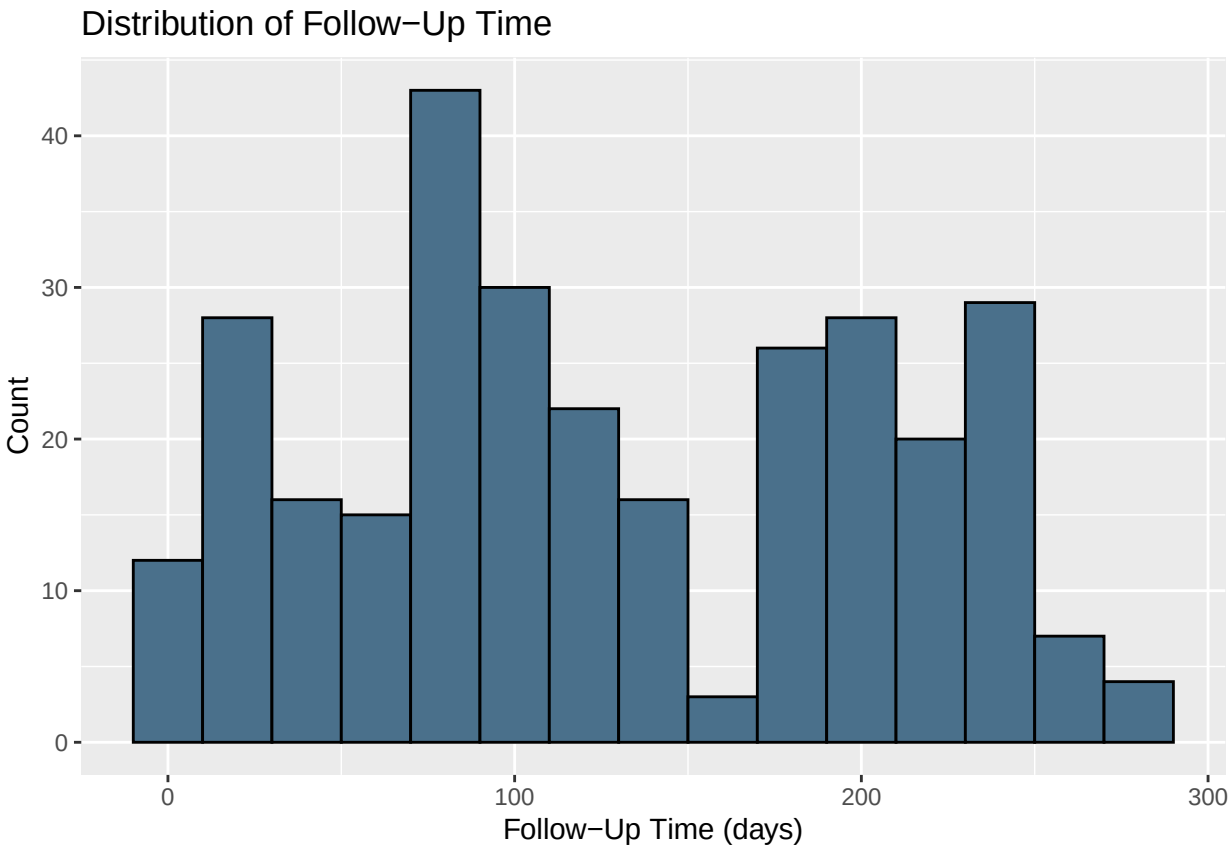
Identifying Relationships with Follow-Up Status [i.e. DEATH_EVENT] for Survival Analysis

A correlation matrix helps identify the strength of association between each variables. In this case, we are more interested in looking at relationships with DEATH_EVENT. Darker shades of red indicate strong negative association and darker shades of blue indicate strong positive association.

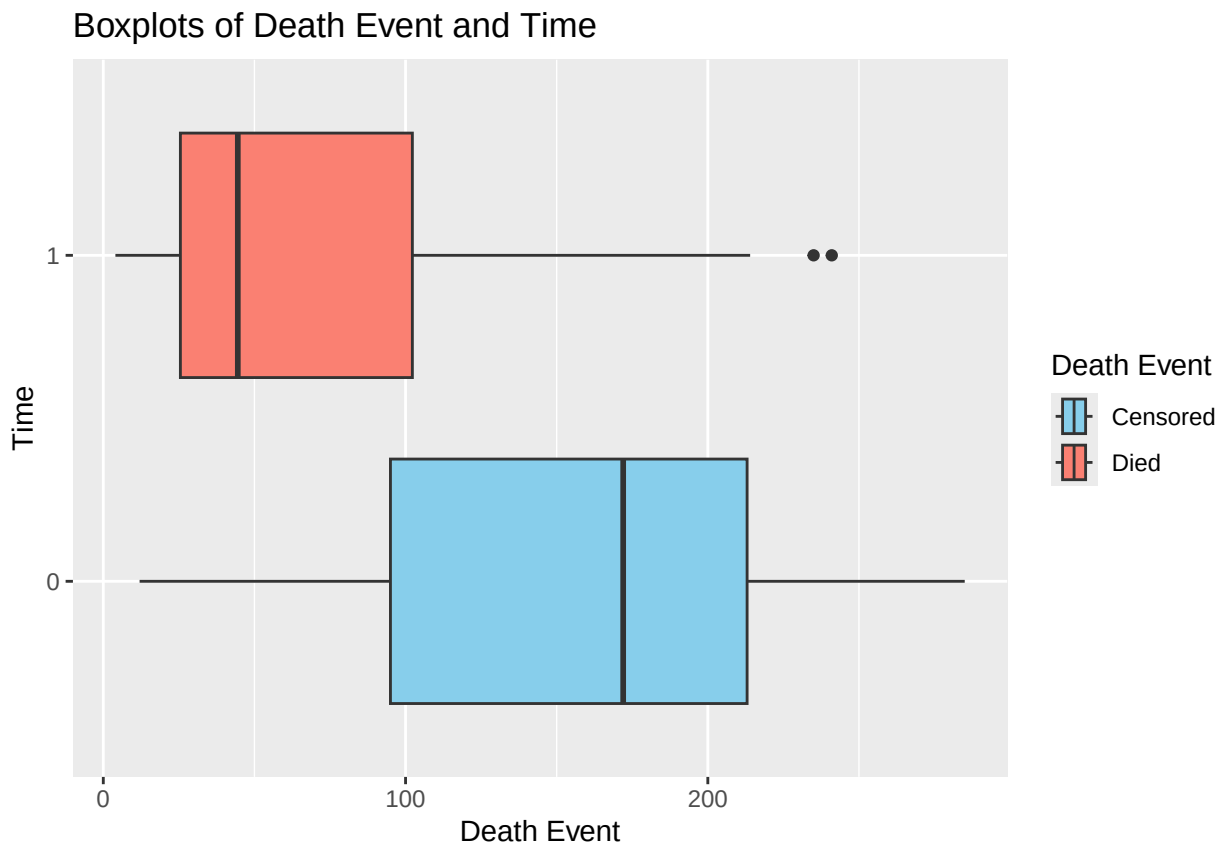


One of the main purposes of this research project is to determine how long a patient can survive with heart failure. In this correlation matrix, we want to understand if a predictor can identify whether a patient will die (i.e. DEATH_EVENT = 1) or not. It seems that the follow-up time is more negatively associated with death event. This further shows that survival analysis would be an appropriate analysis.

Let's take a look at the patients' follow-up times.

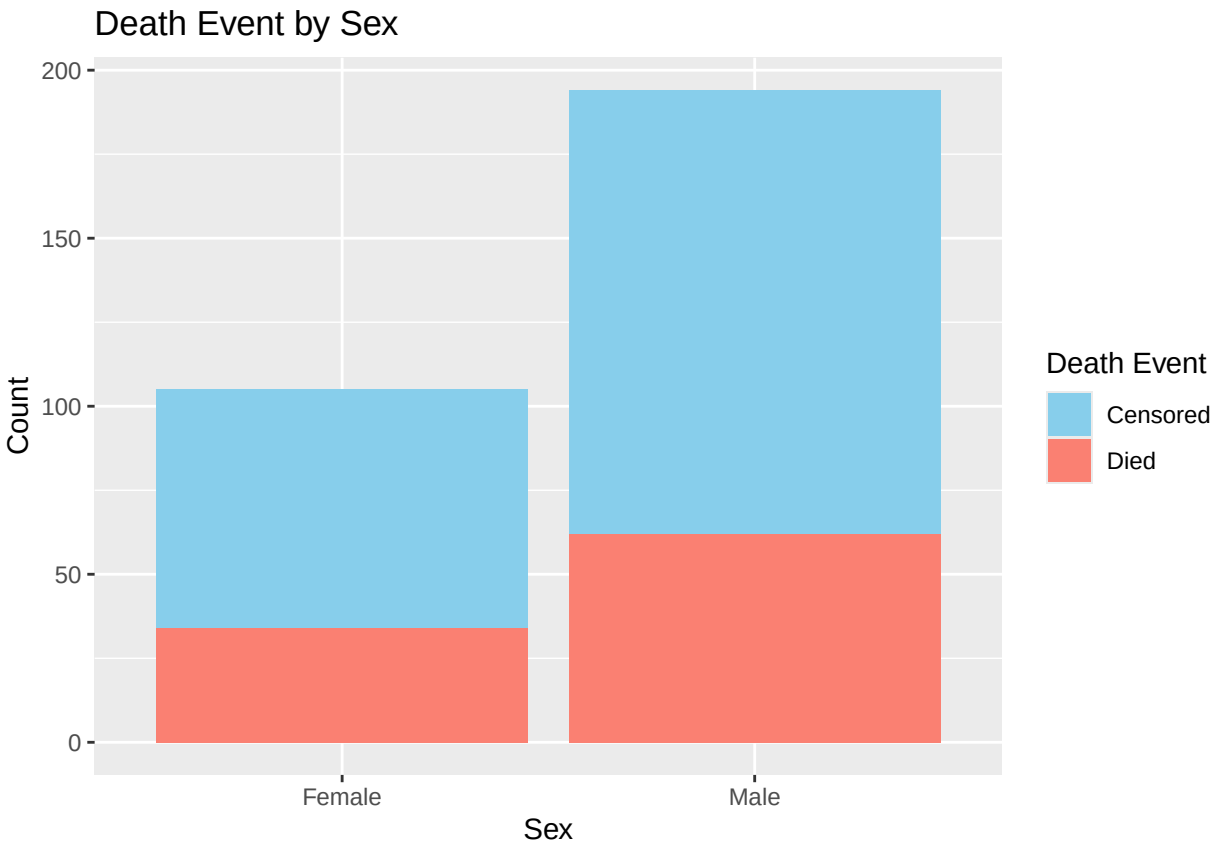


The follow-up time for the patients sees a multimodal distribution that peeks around 20-30 days, 80-90 days, and 200-250 days.



Patients that had died (i.e. DEATH_EVENT = 1) during the follow-up period saw a right-skewed distribution with a median of follow-up time around 50 days. Conversely, patients that were considered "Censored" saw a left-skewed distribution with a median follow-up time of approximately 175 days.

We are also curious to identify DEATH_EVENT by sex.



The death event is similar to both sex, about a quarter of females and males died during the follow-up period.

Grouped by follow-up time in days, the death proportion is calculated. "1" indicates that 100% of patients had died in during that particular time and "0" indicates that no patient died during that time. For example, 50% of patients had died at 29 days of follow-up.

Table 6: Proportion of Death by Follow-Up Time from Days 20 to 30

Time (in Days)	Proportion of Death
15	1
16	0
20	1
22	0
23	1
24	1
26	1
27	1
28	1
29	0.5
30	0.8

Table 7: Follow-Up Time Descriptive Statistics

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
4	73	115	130.3	203	285

Table 8: Death Event Status

Censored	Died
203	96

- The follow-up time is averaged at around 130 days, with the longest being 285 days and the shortest being 4 days.
- 96 out of 299 patients have died during the follow-up period.

Kaplan Meier Method

The Kaplan-Meier method is the most common way to estimate survival times and probabilities. It is a nonparametric maximum likelihood estimate (MLE) for the survival function. We should see a step function, where there is a step down each time a death event occurs.

The `Suv()` function, from the survival package, is to create a survival object from the follow-up time and event variables. This will then be used as a response variable in the model, which will be fitted using the KM estimates. The below code chunk is the survival object with the first thirty survival times. A couple of these survival times are followed by a '+', denoting if the observation is right-censored. Right-censoring is when the event [i.e. death] had not occurred for some patients by the time they were last observed or when the study ended.

```
# use Suv() syntax for right-censored data
survobj <- Surv(time = heart_failure_records$time,
                event = heart_failure_records$DEATH_EVENT
                )
head(survobj, 30)
```

```
##  [1]  4   6   7   7   8   8  10  10  10  10  10  10  11  11  12+ 13  14  14  15
## [20] 15  16+ 20  20  22+ 23  23  24  26  26  26
```

The `survfit()` function produces the survfit object, which fits the default calculations for the KM estimates of the survival curve. These are a step function that jumps at observed death event times. The summary of the survfit object outputs a table for each time of the follow-up time where death occurs. Listed here:

- `n.risk` - the number of patients who were at risk of developing the event
- `n.event` - those who did die
- `survival` - the probability of not dying, or of surviving past that specific time
- standard error and confidence intervals for that probability

```
# "~ 1" signifies that we run the model for overall survival
heart_failure_records_fit <- survfit(survobj ~ 1, data = heart_failure_records)
# we can also print its summary at specific times
summary(heart_failure_records_fit, times =  c(10,50,75,100,150,200))
```

```
## Call: survfit(formula = survobj ~ 1, data = heart_failure_records)
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    10    293      12    0.960  0.0114        0.938        0.982
##    50    244      38    0.831  0.0218        0.789        0.875
##    75    219      13    0.786  0.0239        0.740        0.834
##   100    173       9    0.750  0.0257        0.701        0.802
##   150    118       9    0.703  0.0285        0.649        0.761
##   200     78      11    0.631  0.0330        0.569        0.699
```

As as the number of follow-up days passed, more patients were at risk of dying from heart failure and the probability fo survival had declined. However, the number of patients that did die varied.

Other than the `summary()`function, we can also use the `str()` function to see more details on the structure of the `survfit()` object.

```
# print heart_failure_records_fit object with mean survival time and its standard error
print(heart_failure_records_fit, print.rmean = TRUE)
```

```
## Call: survfit(formula = survobj ~ 1, data = heart_failure_records)
##
##        n events rmean* se(rmean) median 0.95LCL 0.95UCL
## [1,] 299     96    206       6.6     NA      NA      NA
##     * restricted mean with upper limit =  285
```

```
# give more details on the structure of the survfit() object
str(heart_failure_records_fit)
```

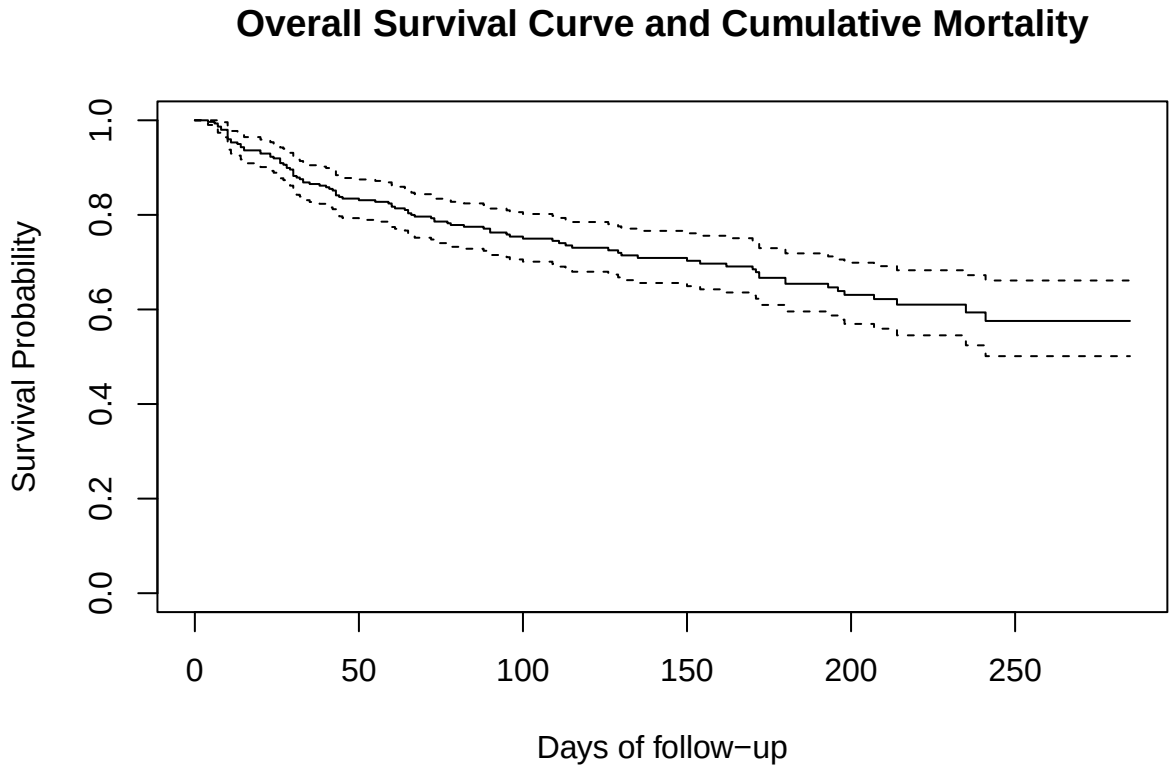
```
## List of 16
##  $ n        : int 299
##  $ time     : num [1:148] 4 6 7 8 10 11 12 13 14 15 ...
##  $ n.risk   : num [1:148] 299 298 297 295 293 287 285 284 283 281 ...
##  $ n.event  : num [1:148] 1 1 2 2 6 2 0 1 2 2 ...
##  $ n.censor : num [1:148] 0 0 0 0 0 0 1 0 0 0 ...
##  $ surv     : num [1:148] 0.997 0.993 0.987 0.98 0.96 ...
##  $ std.err  : num [1:148] 0.00335 0.00475 0.00673 0.00828 0.01183 ...
##  $ cumhaz   : num [1:148] 0.00334 0.0067 0.01343 0.02021 0.04069 ...
##  $ std.chaz : num [1:148] 0.00334 0.00474 0.00672 0.00825 0.01175 ...
##  $ type     : chr "right"
##  $ logse    : logi TRUE
##  $ conf.int : num 0.95
##  $ conf.type: chr "log"
##  $ lower    : num [1:148] 0.99 0.984 0.974 0.964 0.938 ...
##  $ upper    : num [1:148] 1 1 1 0.996 0.982 ...
##  $ call     : language survfit(formula = survobj ~ 1, data = heart_failure_records)
##  - attr(*, "class")= chr "survfit"
```

From the 299 patients, 96 have died from heart failure, with the mean survival time (denoted by `rmean*`) of 206 days and a standard error of 6.6 days.

Plotting Kaplan-Meir Curves

Now that we have fitted the KM estimates, we can visualize the survival probability by follow-up time. All patients were alive at time 0 but probability decreased over time as death occurs. The proportion of patients surviving past 250 days of follow-up is around 60%.

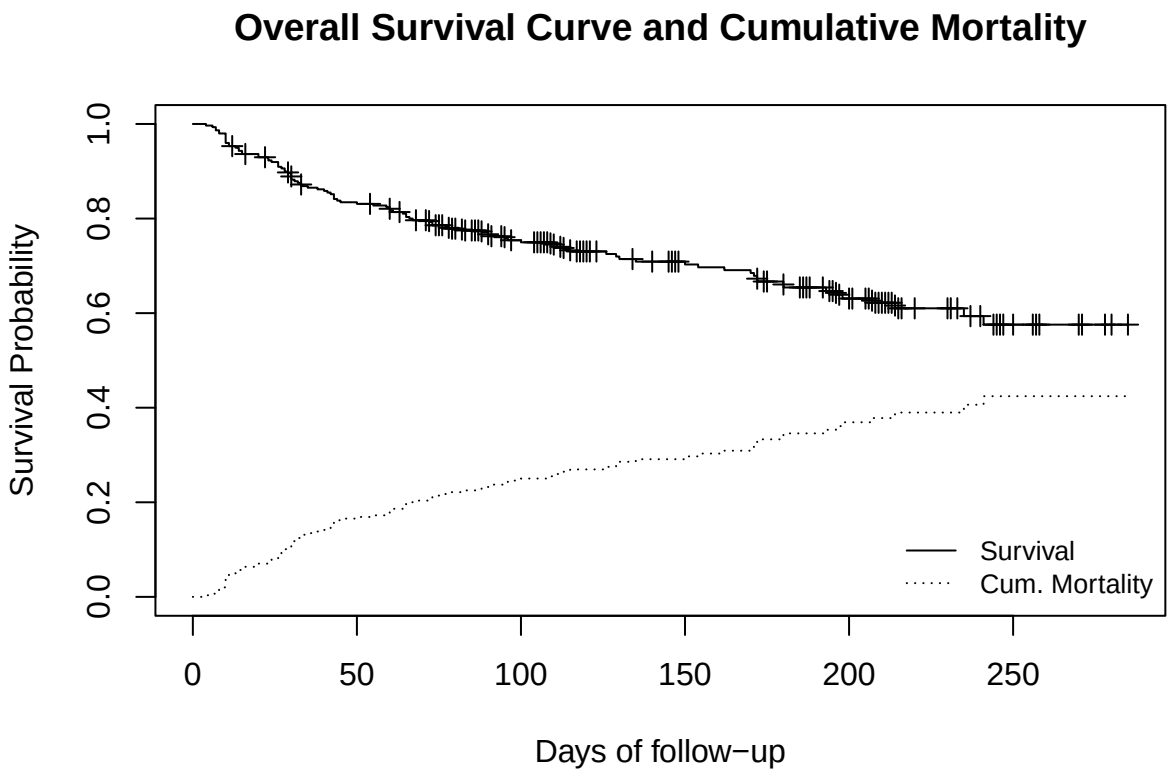
```r
# plot KM survival curve
plot(heart_failure_records_fit,
     xlab = "Days of follow-up",
     ylab = "Survival Probability",
     main = "Overall Survival Curve and Cumulative Mortality"
)
```




```r
# plot survival curve with mark events on the curve: a "+" is printed at every event, death occurence
plot(heart_failure_records_fit,
     xlab = "Days of follow-up",
     ylab = "Survival Probability",
     main = "Overall Survival Curve and Cumulative Mortality",
     mark.time = TRUE,
     conf.int = FALSE
)

# draw additional curve to the previous plot
lines(
  heart_failure_records_fit,
  lty = 3,                        # use different line type for clarity
  fun = "event",                  # draw the cumulative events instead of survival
  mark.time = FALSE,
  conf.int = FALSE
)
```

```r
# add legend to the plot
legend(
  "bottomright",
  legend = c("Survival", "Cum. Mortality"),
  lty = c(1, 3),
  cex = .85,
  bty = "n"
)
```



Log Rank Test

We can also compare survival curves within different groups, visualize their respective survival curves, and run tests to evaluate the difference between independent groups. We will take a look at sex.

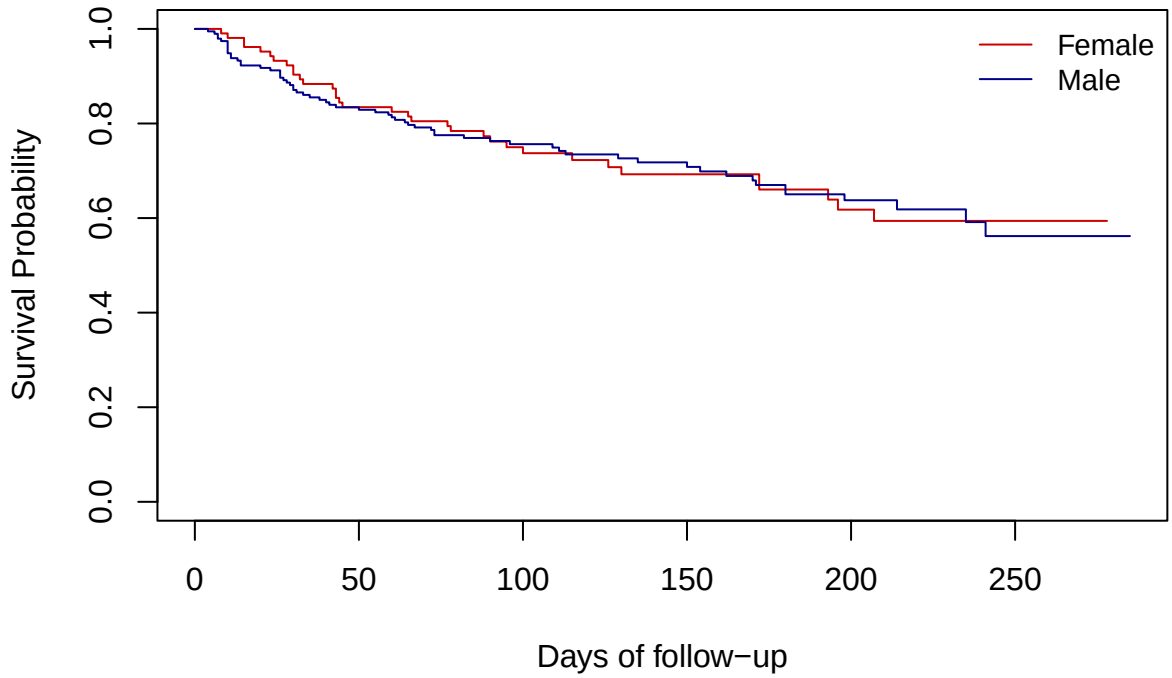
The Log Rank Test is a well-known statistical test that is used to compare the overall survival experiences among two or more independent groups. It evaluates whether the survival curves are identical (overlapping) or not. When survival curves overlap, it suggests that there is not a clear difference in survival experiences between the groups.

```r
# create the new survfit object based on gender
surv_fit_sex <-  survfit(Surv(time, DEATH_EVENT) ~ sex, data = heart_failure_records)

# set colors
col_sex <- c("red3", "blue4")
```

```r
# create plot
plot(
  surv_fit_sex,
  col = col_sex,
  xlab = "Days of follow-up",
  ylab = "Survival Probability")

# add legend
legend(
  "topright",
  legend = c("Female","Male"),
  col = col_sex,
  lty = 1,
  cex = .9,
  bty = "n")
```



The `survdiff()` function of the survival package provides a chi-square statistic along with a p-value.

```r
# compute the test of the difference between survival curves by sex
surv_diff_fit <- survdiff(Surv(time, DEATH_EVENT) ~ sex, data = heart_failure_records)
surv_diff_fit
```

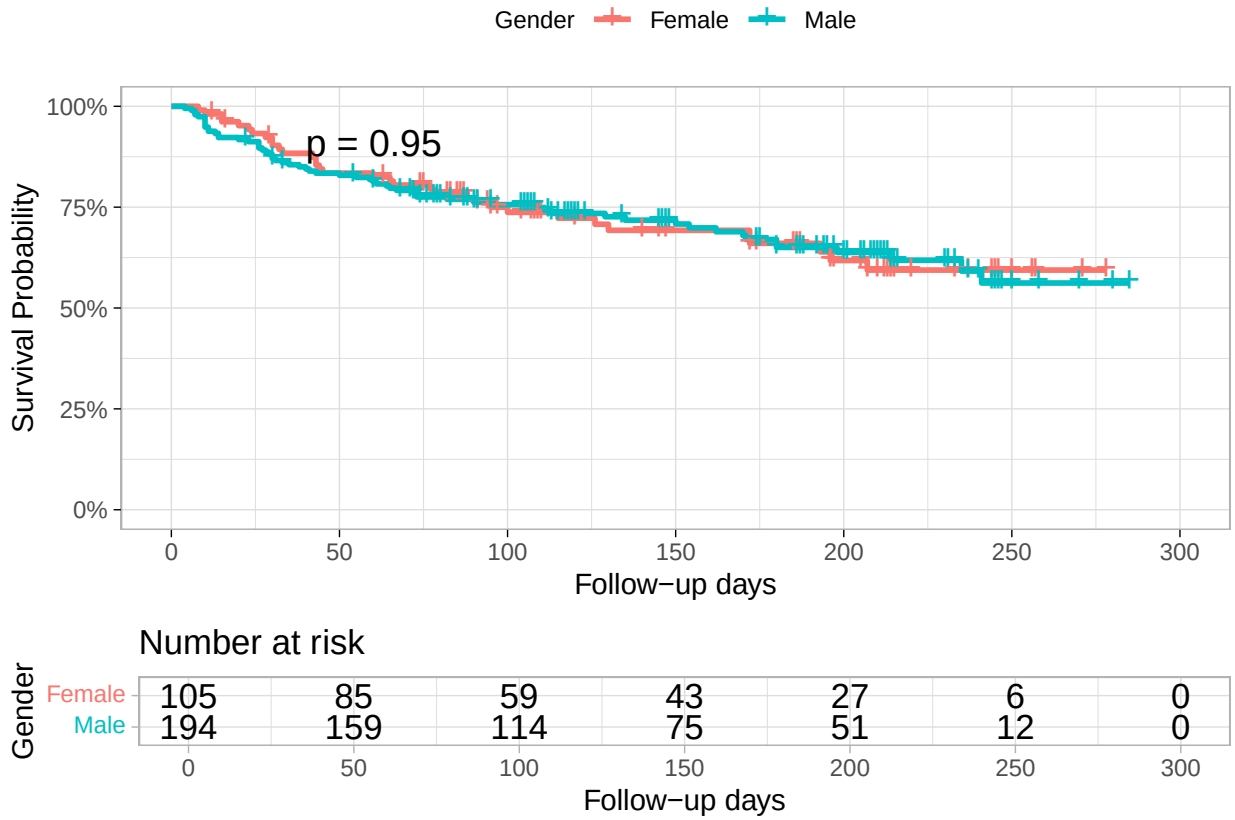
```
## Call:
## survdiff(formula = Surv(time, DEATH_EVENT) ~ sex, data = heart_failure_records)
##
```

```
##         N Observed Expected (O-E)^2/E (O-E)^2/V
## sex=0 105       34     34.3   0.00254   0.00397
## sex=1 194       62     61.7   0.00141   0.00397
##
##  Chisq= 0  on 1 degrees of freedom, p= 0.9
```

The survival curve for females and survival curve for males do overlap and the p-value is at 0.9. Based on this, we can conclude that the log-rank test does not provide enough evidence of a survival difference between females and males.

The `ggsurvplot()` function from the survminer package is a way to illustrate survival curves and test differences all at once.

```
survminer::ggsurvplot(
    surv_fit_sex,
    data = heart_failure_records,
    conf.int = FALSE,              # do not show confidence interval of KM estimates
    surv.scale = "percent",        # present probabilities in the y axis in %
    break.time.by = 50,            # present the time axis with an increment of 10 days
    xlab = "Follow-up days",
    ylab = "Survival Probability",
    pval = T,
    pval.coord = c(40,.91),
    risk.table = T,
    legend.title = "Gender",
    legend.labs = c("Female","Male"),
    font.legend = 9,
    palette = "Dark1",
    surv.median.line = "hv",       # draw horizontal and vertical lines to the median survivals
    ggtheme = theme_light()
)
```



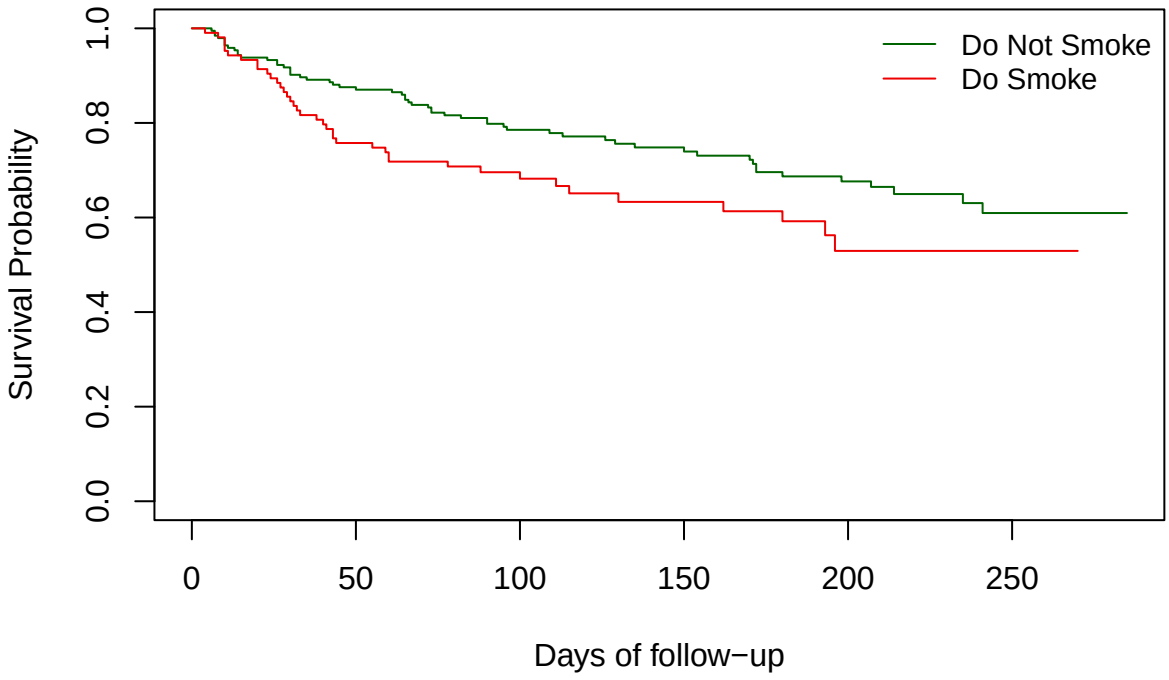
Log Rank Test: Another Example

```
# create the new survfit object based on smoking status
surv_fit_High_blood_pressure <-  survfit(Surv(time, DEATH_EVENT) ~ high_blood_pressure, data = heart_fa

# set colors
col_smoking <- c("darkgreen", "red2")

# create plot
plot(
  surv_fit_High_blood_pressure,
  col = col_smoking,
  xlab = "Days of follow-up",
  ylab = "Survival Probability")

# add legend
legend(
  "topright",
  legend = c("Do Not Smoke","Do Smoke"),
  col = col_smoking,
  lty = 1,
  cex = .9,
  bty = "n")
```




```
surv_diff_fit2 <- survdiff(Surv(time, DEATH_EVENT) ~ high_blood_pressure, data = heart_failure_records)
surv_diff_fit2
```

```
## Call:
## survdiff(formula = Surv(time, DEATH_EVENT) ~ high_blood_pressure,
##     data = heart_failure_records)
##
##                          N Observed Expected (O-E)^2/E (O-E)^2/V
## high_blood_pressure=0 194       57     66.4      1.34      4.41
## high_blood_pressure=1 105       39     29.6      3.00      4.41
##
##  Chisq= 4.4  on 1 degrees of freedom, p= 0.04
```

The survival curve for smokers and survival curve for non-smokers overlap very slightly early on but soon diverges. The p-value is at 0.04, showing that the log-rank test does provide sufficient evidence of a survival difference between non-smokers and smokers.

Cox Regression Analysis

Cox proportional hazards regression is a very popular regression technique for survival analysis. We look at the hazard rate (HR), which measures the risk of failure (i.e. death), given that the patient has survived up to a specific time. It is best to compare independent groups with respect to their hazards and use a hazard ratio for analysis.

- If the hazard ratio for a predictor is close to 1 then that predictor does not affect survival.

- If the HR is less than 1, then the predictor is protective (i.e., associated with improved survival).
- If the HR is greater than 1, then the predictor is associated with increased risk (or decreased survival).

We can first fit a model to access the effect of age and sex on the survival. The `cox.ph()` function from the survival package is used to fit the model, and the function `cox.ph()` from the survival package should be used to test the proportional hazards assumptions for a Cox regression model fit. The hazard ratio is expressed as `exp(coef)`.

```
# fitting the cox model
linelistsurv_cox <-  survival::coxph(
              Surv(time, DEATH_EVENT) ~ age + sex,
              data = heart_failure_records
              )

# printing the model fitted
linelistsurv_cox
```

```
## Call:
## survival::coxph(formula = Surv(time, DEATH_EVENT) ~ age + sex,
##     data = heart_failure_records)
##
##          coef exp(coef)  se(coef)      z        p
## age  0.042222  1.043126  0.008569  4.928 8.33e-07
## sex -0.012871  0.987211  0.213623 -0.060    0.952
##
## Likelihood ratio test=23.52  on 2 df, p=7.815e-06
## n= 299, number of events= 96
```

For each year increase in age, the hazard of death (i.e. risk of death) increases by 4.22%. There is also a slight decrease in the hazard of death for males compared to females. However, the coefficient is not statistically significant (i.e. p-value = 0.952), indicating that sex may not be associated with survival time after considering age.

Let's verify whether the proportional hazards assumptions is satisfied before moving forward.

```
# verify proportional hazards assumptions
test_ph_ <- survival::cox.zph(linelistsurv_cox)
test_ph_
```

```
##        chisq df    p
## age    0.791  1 0.37
## sex    0.315  1 0.57
## GLOBAL 1.139  2 0.57
```

Since the p-value is greater than 0.05 for age and sex individually, we fail to reject the null hypothesis and the proportional hazards assumption is not violated. There is also no violation of the proportional hazards assumption globally across all predictor variables. Therefore, we can proceed with interpreting the coefficients.

Let's add more risk factors such as ejection_fraction, high blood pressure, serum_creatinine, and smoking. Of course, we still need to verify the proportion hazards assumption before going forward.

```r
# fitting the cox model
linelistsurv_cox2 <-  survival::coxph(
          Surv(time, DEATH_EVENT) ~ age + sex + high_blood_pressure + serum_creatinine + smoking,
          data = heart_failure_records
          )

# printing the model fitted
linelistsurv_cox2
```

```
## Call:
## survival::coxph(formula = Surv(time, DEATH_EVENT) ~ age + sex +
##     high_blood_pressure + serum_creatinine + smoking, data = heart_failure_records)
##
##                          coef exp(coef) se(coef)     z        p
## age                  0.042547  1.043465 0.008873 4.795 1.63e-06
## sex                  0.037060  1.037755 0.242425 0.153   0.8785
## high_blood_pressure 0.390670  1.477971 0.212874 1.835   0.0665
## serum_creatinine    0.287639  1.333276 0.057869 4.970 6.68e-07
## smoking              0.119772  1.127239 0.247658 0.484   0.6287
##
## Likelihood ratio test=43.99  on 5 df, p=2.323e-08
## n= 299, number of events= 96
```

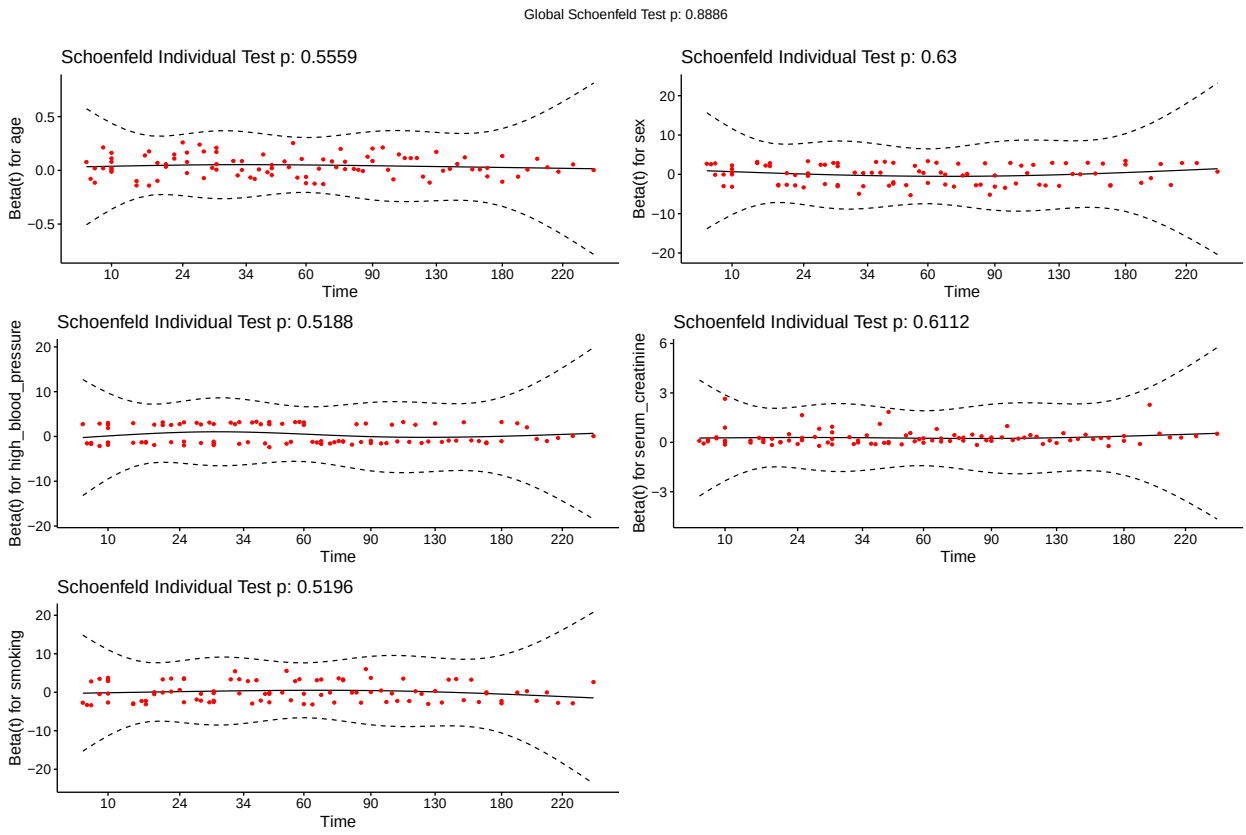
```r
# test the proportional hazard model
linelistsurv_test <- cox.zph(linelistsurv_cox2)
linelistsurv_test
```

```
##                      chisq df    p
## age                  0.347  1 0.56
## sex                  0.232  1 0.63
## high_blood_pressure 0.416  1 0.52
## serum_creatinine    0.258  1 0.61
## smoking              0.415  1 0.52
## GLOBAL               1.702  5 0.89
```

All predictors seem to be significanntly associated with survival time except for sex and smoking status.

We can also graphically verify this proportion hazards assumption using `ggcoszph()` from the survminer package.

```r
survminer::ggcoxzph(linelistsurv_test)
```

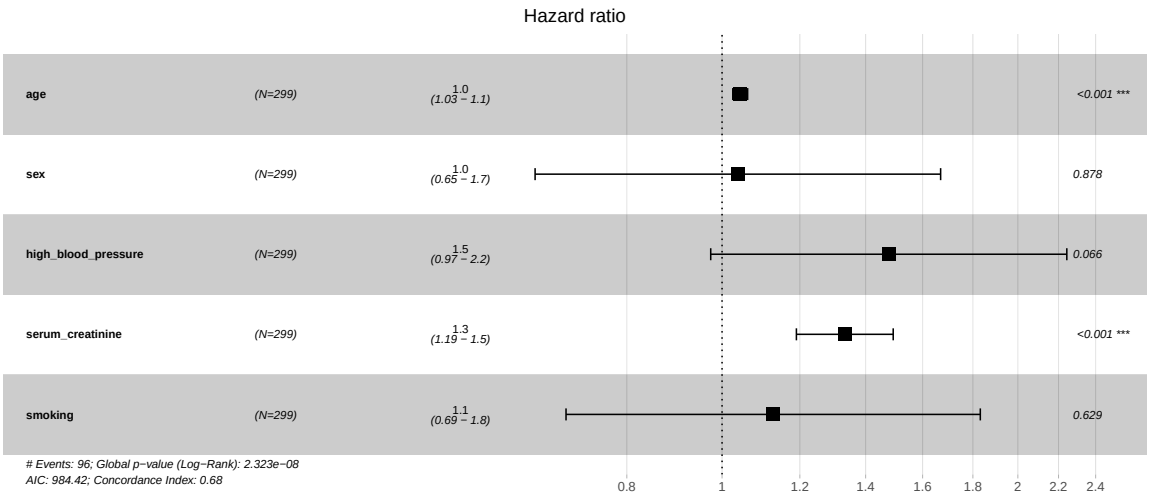



```
## Call:
## survival::coxph(formula = Surv(time, DEATH_EVENT) ~ age + sex +
##     high_blood_pressure + serum_creatinine + smoking, data = heart_failure_records)
##
##   n= 299, number of events= 96
##
##                        coef exp(coef) se(coef)     z Pr(>|z|)
## age                0.042547  1.043465 0.008873 4.795 1.63e-06 ***
## sex                0.037060  1.037755 0.242425 0.153   0.8785
## high_blood_pressure 0.390670  1.477971 0.212874 1.835   0.0665 .
## serum_creatinine    0.287639  1.333276 0.057869 4.970 6.68e-07 ***
## smoking            0.119772  1.127239 0.247658 0.484   0.6287
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                    exp(coef) exp(-coef) lower .95 upper .95
## age                    1.043     0.9583    1.0255     1.062
## sex                    1.038     0.9636    0.6453     1.669
## high_blood_pressure    1.478     0.6766    0.9738     2.243
## serum_creatinine       1.333     0.7500    1.1903     1.493
## smoking                1.127     0.8871    0.6938     1.832
##
## Concordance= 0.675  (se = 0.03 )
## Likelihood ratio test= 43.99  on 5 df,   p=2e-08
## Wald test            = 50.98  on 5 df,   p=9e-10
## Score (logrank) test = 55.63  on 5 df,   p=1e-10
```

Although the hazard ratios for age, high blood pressure, and the serum creatinine are fairly close to 1, the p-values are still significant, suggesting that these predictors can be lightly associated with increased risk of death.

We can visualize the results of the cox model using the forest plots with the `ggforest()` function from the survminer package.

```
# visual results of cox model
ggforest(linelistsurv_cox2, data = heart_failure_records)
```



Results

We investigated on which clinical features (i.e. health conditions and diagnoses) that may be evident in heart failure. Creatinine phophokinase (CPK) levels and the ejection fraction are possible indicators of heart failure. A significant portion of patients with heart failure in this study had CPK levels beyond the normal range and low ejection fractions that fell below the healthy range. High CPK levels showed stress in the heart and low ejection fraction showed the the heart's inability to pump blood.

From our survival analyses, we were also able to better capture the changes in the survival of heart failure over time. Through the Kaplan Meier method, we fitted KM estimates to approximate survival times and probabilities. Based on the KM curves, the probability of survival gradually declined after each follow-up day, leveling at around 60% after 250 days.

Through the log rank test, we compared survival curves by sex, however, there was not enough evidence for survival difference between females and males. Conversely, there was a survival difference between non-smokers and smokers, with non-smokers having a higher probability of survival.

Finally, we took a step further through the cox regression analysis to measure the risk of death, from the hazard ratio, to pinpoint which predictors have an effect on survival. The age of the patient, whether or not the patient had high blood pressure, and the serum creatinine level all have a significant effect of the the survival, suggesting that those predictors can have an effect on the increased risk of death.

For future analysis, it is important to select a large-enough dataset that has both a time and event variable to properly conduct survival analysis. It is also crucial to test the proportional hazards assumptions for a Cox regression model fit before interpreting the model so no assumptions are violated. Violations could lead to biased estimates and other inaccuracies. The hazard ratio between any two groups should remain constant over time.

Bibliography

@misc{Batra, Neale, et al._2023, title={The Epidemiologist R Handbook}, url={https://epirhandbook.com/en/survival-analysis.html}, journal={27 Survival analysis}, author={Batra, Neale, et al.}, year={2023}, month={Jul}}

@misc{Centers for Disease Control and Prevention_2023, url={https://www.cdc.gov/heartdisease/heart_failure.htm}, journal={Centers for Disease Control and Prevention}, publisher={Centers for Disease Control and Prevention}, year={2023}, month={Jan}}

@misc{Mayo Clinic_2024, url={https://www.mayoclinic.org/diseases-conditions}, journal={Mayo Clinic}, publisher={Mayo Foundation for Medical Education and Research}, year={2024}}

@misc{misc_heart_failure_clinical_records_519, title = {{Heart Failure Clinical Records}}, year = {2020}, howpublished = {UCI Machine Learning Repository}, note = {{DOI}: https://doi.org/10.24432/C5Z89R} }

@misc{Mount Sinai Health System_2024, url={https://www.mountsinai.org/health-library/tests/creatine-phosphokinase-test}, journal={Mount Sinai Health System}, year={2024}}