```r
heart_failure_data$high_blood_pressure <- as.factor(heart_failure_data$high_blood_pressure)
heart_failure_data$anaemia <- as.factor(heart_failure_data$anaemia)
heart_failure_data$diabetes <- as.factor(heart_failure_data$diabetes)
heart_failure_data$sex <- as.factor(heart_failure_data$sex)

heart_failure_data <- heart_failure_data |>
  mutate(
    log_cp = log(creatinine_phosphokinase),
    log_serum_creatinine = log(serum_creatinine)
  )

## COX
### select covariates with alpha = 0.2
coxph(
  Surv(time, DEATH_EVENT) ~ age,
  data = heart_failure_data
)

coxph(
  Surv(time, DEATH_EVENT) ~ log_cp,
  data = heart_failure_data
)

coxph(
  Surv(time, DEATH_EVENT) ~ ejection_fraction,
  data = heart_failure_data
)

coxph(
  Surv(time, DEATH_EVENT) ~ platelets,
  data = heart_failure_data
)

coxph(
  Surv(time, DEATH_EVENT) ~ anaemia,
  data = heart_failure_data
)

coxph(
  Surv(time, DEATH_EVENT) ~ log_serum_creatinine,
  data = heart_failure_data
)

coxph(
  Surv(time, DEATH_EVENT) ~ serum_sodium,
  data = heart_failure_data
)

coxph(
  Surv(time, DEATH_EVENT) ~ diabetes,
  data = heart_failure_data
)
```

```r
coxph(
  Surv(time, DEATH_EVENT) ~ high_blood_pressure,
  data = heart_failure_data
)

coxph(
  Surv(time, DEATH_EVENT) ~ sex,
  data = heart_failure_data
)

coxph(
  Surv(time, DEATH_EVENT) ~ smoking,
  data = heart_failure_data
)

# age ejection_fraction anaemia log_serum_creatinine serum_sodium high_blood_pressure

# Define the Cox proportional hazards model
cox_full <- coxph(
  Surv(time, DEATH_EVENT) ~ age + ejection_fraction + log_serum_creatinine + serum_sodium + anaemia + hi
  data = heart_failure_data
)

cox_backward <- step(cox_full, direction = "backward")
# out:serum_sodium
summary(cox_backward)

#forward stepwise
cox_null <- coxph(Surv(time, DEATH_EVENT) ~ 1, data = heart_failure_data)
cox_forward <- step(cox_null,
                    scope = ~ age + log_cp + ejection_fraction +
                      platelets + log_serum_creatinine + serum_sodium +
                      anaemia + diabetes + high_blood_pressure + sex + smoking,
                    direction = "forward")
summary(cox_forward)

# interactions

cox_interaction <- coxph(
  Surv(time, DEATH_EVENT) ~ age +
    high_blood_pressure +
    anaemia +
    ejection_fraction +
    log_serum_creatinine +
    age * high_blood_pressure +
    age * anaemia +
    age * ejection_fraction +
    age * log_serum_creatinine +
    high_blood_pressure * anaemia +
    high_blood_pressure * ejection_fraction +
    high_blood_pressure * log_serum_creatinine +
    anaemia * ejection_fraction +
    anaemia * log_serum_creatinine +
```

```
    ejection_fraction * log_serum_creatinine,
  data = heart_failure_data
)

cox_int <- stepAIC(
  cox_interaction,
  direction = "both",
  trace = TRUE
)

summary(cox_int)

aic_comparison <- AIC(cox_full, cox_backward, cox_forward, cox_int)
print(aic_comparison)

# VIF test for final selected variables
vif(lm(time ~ age + ejection_fraction + log_serum_creatinine +
       anaemia + high_blood_pressure + age*ejection_fraction + ejection_fraction*log_serum_creatinine,
       data = heart_failure_data),type = 'predictor')
```

```
## GVIFs computed for predictors
```

```
## Time-varying

### selected model with time-varying variable ejection_fraction
heart_failure_data$log_time <- log(heart_failure_data$time)

cox_time <- coxph(
  Surv(time, DEATH_EVENT) ~ age +
    high_blood_pressure +
    anaemia +
    #ejection_fraction +
    log_serum_creatinine +
    age * ejection_fraction +
    ejection_fraction * log_time +
    ejection_fraction * log_serum_creatinine,
  data = heart_failure_data
)
```

```
## Warning in coxph.fit(X, Y, istrat, offset, init, control, weights = weights, :
## Ran out of iterations and did not converge

## Warning in coxph.fit(X, Y, istrat, offset, init, control, weights = weights, :
## one or more coefficients may be infinite
```

```
summary(cox_time)

AIC(cox_time)

# piecewise cox model

heart_failure_data$time_group <- cut(heart_failure_data$time,
```

```
                                    breaks=c(0, 100, 200, Inf),
                                    labels=c("0-100", "100-200", ">200"))


cox_piecewise <- coxph(
  Surv(time, DEATH_EVENT) ~ age + high_blood_pressure + anaemia +
  log_serum_creatinine + ejection_fraction * time_group,
  data = heart_failure_data
)
```

```
## Warning in coxph.fit(X, Y, istrat, offset, init, control, weights = weights, :
## Loglik converged before variable 6,7 ; coefficient may be infinite.
```

```
## AFT
### weibull step 1
survreg(
  Surv(time, DEATH_EVENT) ~ age,
  dist = "weibull",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ log_cp,
  dist = "weibull",
  data = heart_failure_data
)


survreg(
  Surv(time, DEATH_EVENT) ~ ejection_fraction,
  dist = "weibull",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ platelets,
  dist = "weibull",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ log_serum_creatinine,
  dist = "weibull",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ serum_sodium,
  dist = "weibull",
  data = heart_failure_data
)

survreg(
```

```
  Surv(time, DEATH_EVENT) ~ anaemia,
  dist = "weibull",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ diabetes,
  dist = "weibull",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ high_blood_pressure,
  dist = "weibull",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ sex,
  dist = "weibull",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ smoking,
  dist = "weibull",
  data = heart_failure_data
)

# high_blood_pressure anaemia serum_sodium log_serum_creatinine ejection_fraction age

### weibull step 2
weibull_full <-
survreg(
  Surv(time, DEATH_EVENT) ~ age + ejection_fraction + log_serum_creatinine + serum_sodium + anaemia + hi
  dist = "weibull",
  data = heart_failure_data
)

weib_backward <- step(weibull_full, direction = "backward")
summary(weib_backward)
# out:serum_sodium


#weibull step 3
weibull_null <- survreg(Surv(time, DEATH_EVENT) ~ age+ejection_fraction+log_serum_creatinine+anaemia+hig
weib_forward <- step(weibull_null,
                     scope = ~ age + log_cp + ejection_fraction +
                       platelets + log_serum_creatinine + serum_sodium +
                       anaemia + diabetes + high_blood_pressure + sex + smoking,
,data = heart_failure_data,
                     direction = "forward")
```

```r
summary(weib_forward)

### Weibull step 4
weib_interaction <- survreg(
  Surv(time, DEATH_EVENT) ~ age +
    high_blood_pressure +
    anaemia +
    ejection_fraction +
    log_serum_creatinine +
    age * high_blood_pressure +
    age * anaemia +
    age * ejection_fraction +
    age * log_serum_creatinine +
    high_blood_pressure * anaemia +
    high_blood_pressure * ejection_fraction +
    high_blood_pressure * log_serum_creatinine +
    anaemia * ejection_fraction +
    anaemia * log_serum_creatinine +
    ejection_fraction * log_serum_creatinine,
  dist = "weibull",
  data = heart_failure_data
)
weib_int <- stepAIC(weib_interaction, direction = "both")
summary(weib_int)

### exponential 1
survreg(
  Surv(time, DEATH_EVENT) ~ age,
  dist = "exponential",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ log_cp,
  dist = "exponential",
  data = heart_failure_data
)


survreg(
  Surv(time, DEATH_EVENT) ~ ejection_fraction,
  dist = "exponential",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ platelets,
  dist = "exponential",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ log_serum_creatinine,
```

```
  dist = "exponential",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ serum_sodium,
  dist = "exponential",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ anaemia,
  dist = "exponential",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ diabetes,
  dist = "exponential",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ high_blood_pressure,
  dist = "exponential",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ sex,
  dist = "exponential",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ smoking,
  dist = "exponential",
  data = heart_failure_data
)

# high_blood_pressure anaemia serum_sodium log_serum_creatinine ejection_fraction age

### exponential 2
exponential_full <-
survreg(
  Surv(time, DEATH_EVENT) ~ age + ejection_fraction + log_serum_creatinine + serum_sodium + anaemia + hi
  dist = "exponential",
  data = heart_failure_data
)

exp_backward <- step(exponential_full, direction = "backward")
summary(exp_backward)
# out:serum_sodium
```

```r
### exponential step 3
exponential_null <- survreg(Surv(time, DEATH_EVENT) ~  age+ejection_fraction+log_serum_creatinine+anaem
exp_forward <- step(exponential_null,
                    scope = ~ age + log_cp + ejection_fraction +
                      platelets + log_serum_creatinine + serum_sodium +
                      anaemia + diabetes + high_blood_pressure + sex + smoking,
,data = heart_failure_data,
                    direction = "forward")

summary(exp_forward)

### exponential step 4
exp_interaction <- survreg(
  Surv(time, DEATH_EVENT) ~ age +
    high_blood_pressure +
    anaemia +
    ejection_fraction +
    log_serum_creatinine +
    age * high_blood_pressure +
    age * anaemia +
    age * ejection_fraction +
    age * log_serum_creatinine +
    high_blood_pressure * anaemia +
    high_blood_pressure * ejection_fraction +
    high_blood_pressure * log_serum_creatinine +
    anaemia * ejection_fraction +
    anaemia * log_serum_creatinine +
    ejection_fraction * log_serum_creatinine,
  dist = "exponential",
  data = heart_failure_data
)
exp_int <- stepAIC(exp_interaction, direction = "both")
summary(exp_int)

### log-logistic step 1
survreg(
  Surv(time, DEATH_EVENT) ~ age,
  dist = "loglogistic",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ log_cp,
  dist = "loglogistic",
  data = heart_failure_data
)


survreg(
  Surv(time, DEATH_EVENT) ~ ejection_fraction,
  dist = "loglogistic",
  data = heart_failure_data
)
```

```r
survreg(
  Surv(time, DEATH_EVENT) ~ platelets,
  dist = "loglogistic",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ log_serum_creatinine,
  dist = "loglogistic",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ serum_sodium,
  dist = "loglogistic",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ anaemia,
  dist = "loglogistic",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ diabetes,
  dist = "loglogistic",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ high_blood_pressure,
  dist = "loglogistic",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ sex,
  dist = "loglogistic",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ smoking,
  dist = "loglogistic",
  data = heart_failure_data
)

# high_blood_pressure anaemia serum_sodium log_serum_creatinine ejection_fraction age
### log-logistic step 2
loglogistic_full <-
survreg(
  Surv(time, DEATH_EVENT) ~ age + ejection_fraction + log_serum_creatinine + serum_sodium + anaemia + hi
```

```
  dist = "loglogistic",
  data = heart_failure_data
)

llog_backward <- step(loglogistic_full, direction = "backward")
summary(llog_backward)
# out:serum_sodium

### log-logistic step 3
loglogistic_null <- survreg(Surv(time, DEATH_EVENT) ~ age+anaemia+log_serum_creatinine+high_blood_pressu
llog_forward <- step(loglogistic_null,
                     scope = ~ age + log_cp + ejection_fraction +
                       platelets + log_serum_creatinine + serum_sodium +
                       anaemia + diabetes + high_blood_pressure + sex + smoking,
,data = heart_failure_data,
                     direction = "forward")

summary(llog_forward)

### log-logistic step 4
llog_interaction <- survreg(
  Surv(time, DEATH_EVENT) ~ age +
    high_blood_pressure +
    anaemia +
    ejection_fraction +
    log_serum_creatinine +
    age * high_blood_pressure +
    age * anaemia +
    age * ejection_fraction +
    age * log_serum_creatinine +
    high_blood_pressure * anaemia +
    high_blood_pressure * ejection_fraction +
    high_blood_pressure * log_serum_creatinine +
    anaemia * ejection_fraction +
    anaemia * log_serum_creatinine +
    ejection_fraction * log_serum_creatinine,
  dist = "loglogistic",
  data = heart_failure_data
)
llog_int <- stepAIC(llog_interaction, direction = "both")
summary(llog_int)

### log-normal step 1
survreg(
  Surv(time, DEATH_EVENT) ~ age,
  dist = "lognormal",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ log_cp,
  dist = "lognormal",
  data = heart_failure_data
```

```
)


survreg(
  Surv(time, DEATH_EVENT) ~ ejection_fraction,
  dist = "lognormal",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ platelets,
  dist = "lognormal",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ log_serum_creatinine,
  dist = "lognormal",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ serum_sodium,
  dist = "lognormal",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ anaemia,
  dist = "lognormal",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ diabetes,
  dist = "lognormal",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ high_blood_pressure,
  dist = "lognormal",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ sex,
  dist = "lognormal",
  data = heart_failure_data
)

survreg(
  Surv(time, DEATH_EVENT) ~ smoking,
```

```
  dist = "lognormal",
  data = heart_failure_data
)

# high_blood_pressure anaemia serum_sodium log_serum_creatinine ejection_fraction age

### log-normal step 2
lognormal_full <-
survreg(
  Surv(time, DEATH_EVENT) ~ age + ejection_fraction + log_serum_creatinine + serum_sodium + anaemia + hi
  dist = "lognormal",
  data = heart_failure_data
)

logn_backward <- step(lognormal_full, direction = "backward")
summary(logn_backward)

#step 3
lognormal_null <- survreg(Surv(time, DEATH_EVENT) ~ age+anaemia+log_serum_creatinine+ejection_fraction+l
logn_forward <- step(lognormal_null,
                     scope = ~ age + log_cp + ejection_fraction +
                       platelets + log_serum_creatinine + serum_sodium +
                       anaemia + diabetes + high_blood_pressure + sex + smoking,
,data = heart_failure_data,
                     direction = "forward")

summary(logn_forward)

### log_normal step 4
logn_interaction <- survreg(
  Surv(time, DEATH_EVENT) ~ age +
    high_blood_pressure +
    anaemia +
    ejection_fraction +
    log_serum_creatinine +
    serum_sodium+
    age*serum_sodium+
    serum_sodium*high_blood_pressure+
    serum_sodium*anaemia +
    serum_sodium*ejection_fraction +
    serum_sodium*log_serum_creatinine +
    age * high_blood_pressure +
    age * anaemia +
    age * ejection_fraction +
    age * log_serum_creatinine +
    high_blood_pressure * anaemia +
    high_blood_pressure * ejection_fraction +
    high_blood_pressure * log_serum_creatinine +
    anaemia * ejection_fraction +
    anaemia * log_serum_creatinine +
    ejection_fraction * log_serum_creatinine,
  dist = "lognormal",
  data = heart_failure_data
```

```r
)
logn_int <- stepAIC(logn_interaction, direction = "both")
summary(logn_int)

AIC(exp_int,llog_int,logn_int,cox_int,weib_int)

# expoential model as AFT model

### exponential model check
surv_obj <- Surv(heart_failure_data$time, heart_failure_data$DEATH_EVENT)
fit_total <- survfit(surv_obj ~ 1, data = heart_failure_data)

g_total <- ggsurvplot(
  fit_total,
  data = heart_failure_data,
  conf.int = FALSE,
  xlab = "Time",
  ylab = "Cumulative Hazard",
  ggtheme = theme_minimal(),
  risk.table = FALSE,
  fun = "cumhaz",
  main = "Cumulative Hazard (Total)",
  palette = "black"
)
```