ACCELERATED FAILURE TIME MODELS

Enock Bereka

2025-08-25

Load required modules

```
library(survival)
library(flexsurv)
library(tidyverse)
```

Load the data

```
dat <- read.csv("C:/Users/ADMIN/Desktop/Data
Science/Datasets/survival/exponential_survival_data.csv")
```

Quick check

```
glimpse(dat)
```

```
## Rows: 300
## Columns: 7
## $ id        <int> 263, 115, 166, 279, 294, 170, 12, 236, 241, 207, 195,
224, 5…
## $ time      <dbl> 11.600, 9.230, 6.112, 2.421, 8.198, 6.765, 13.899,
0.706, 5.…
## $ status    <int> 0, 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 1, 1, 0, 0, 0,
0, …
## $ treatment <int> 0, 1, 1, 1, 0, 1, 1, 1, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0,
1, …
## $ sex       <int> 0, 1, 0, 0, 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 0, 0,
0, …
## $ age       <dbl> 56.1, 36.3, 57.2, 59.1, 47.4, 68.5, 52.8, 59.4, 60.9,
49.3, …
## $ biomarker <dbl> -0.651, 0.659, 0.194, 0.399, -1.576, -0.700, -0.703,
0.027, …
```

```
with(dat, table(status))
```

```
## status
##   0   1
## 159 141
```

Accelerates Failure Time

It's a type of survival model where the covariates are assumed to act by accelerating or decelerating the survival time directly, rather than modifying the hazard ratio (like in Cox PH models).

Exponential model via survival::survreg (AFT form)

Note: survreg() with dist="exponential" fits an AFT model.

Coefficients are on log(time) scale (time ratios). Negative coef -> shorter survival (higher hazard).

```
fit_aft <- survreg(Surv(time, status) ~ treatment + sex + age + biomarker,
                   data = dat, dist = "exponential")
summary(fit_aft)
```

```
##
## Call:
## survreg(formula = Surv(time, status) ~ treatment + sex + age +
##     biomarker, data = dat, dist = "exponential")
##                Value Std. Error     z       p
## (Intercept)  3.52516    0.48735  7.23 4.7e-13
## treatment    0.31374    0.17032  1.84   0.065
## sex         -0.20205    0.16881 -1.20   0.231
## age         -0.01590    0.00835 -1.90   0.057
## biomarker   -0.39547    0.09206 -4.30 1.7e-05
##
## Scale fixed at 1
##
## Exponential distribution
## Loglik(model)= -499.4   Loglik(intercept only)= -513.3
##  Chisq= 27.77 on 4 degrees of freedom, p= 1.4e-05
## Number of Newton-Raphson Iterations: 5
## n= 300
```

Model type

You fit an Accelerated Failure Time (AFT) model with exponential distribution.

Coefficients are on the log(time) scale.

Positive coefficients → longer survival (slower hazard).

Negative coefficients → shorter survival (faster hazard).

Intercept (3.53): baseline log(survival time) when all covariates = 0. (Not usually of direct interest.)

Treatment (0.31, p=0.065): Positive effect. Suggests treatment extends survival times by a factor of

exp(0.314) ≈ 1.37.

→ Patients on treatment survive ~37% longer (median) than controls, but borderline significant.

Sex (-0.20, p=0.231): Negative but not significant. Males (coded 1) tend to have ~18% shorter survival (exp(-0.202) ≈ 0.82), but evidence is weak.

Age (-0.016, p=0.057): Each extra year of age decreases survival time by about exp(-0.016) ≈ 0.98 (~2% reduction per year). Marginal significance.

Biomarker (-0.395, $p < 0.001$): Strongly negative. Each +1 SD increase in biomarker reduces survival time by exp(-0.395) ≈ 0.67.

→ That means about 33% shorter survival, highly significant.

Model fit statistics

Loglik(model) = -499.4 vs intercept-only = -513.3

Likelihood ratio test: $Chi^2 = 27.8$, df=4, $p < 0.0001$

→ The covariates together significantly improve the model fit.

Scale fixed at 1 → because exponential distribution has only one parameter (constant hazard).

Summary of practical meaning

Treatment appears beneficial (≈37% longer survival), though borderline significant (p=0.065).

Older age trends toward worse survival (≈2% shorter per year, borderline p=0.057).

Male sex shows worse survival (≈18% shorter), but not significant (p=0.231).

High biomarker levels strongly predict worse survival (≈33% shorter per unit increase, p<0.001).

So the biomarker is the strongest predictor, with treatment showing promising (but borderline) benefit.

Convert AFT coef to approximate hazard ratios (HR ≈ exp(-beta_AFT))

(Exact mapping differs because AFT vs PH, but this gives a handy interpretation.)

```
hr_aft <- exp(-coef(fit_aft))
cbind(HR_approx = hr_aft)
```

```
##              HR_approx
## (Intercept) 0.02944704
## treatment   0.73071081
## sex         1.22391333
## age         1.01602736
## biomarker   1.48508818
```

Treatment (HR ≈ 0.73)

→ Hazard of death is about 27% lower for patients on treatment compared to control.

→ Protective effect, consistent with longer survival. Borderline significant from earlier p=0.065.

Sex (HR ≈ 1.22)

→ Males have ~22% higher hazard compared to females.

→ Matches the negative coefficient in AFT (shorter survival for males), though not statistically significant (p=0.231).

Age (HR ≈ 1.02)

→ Each additional year of age increases hazard by ~2%.

→ Over 10 years, hazard is ~22% higher (1.02^10 ≈ 1.22). Borderline significant (p=0.057).

Biomarker (HR ≈ 1.49)

→ Each 1-unit increase in biomarker raises hazard by ~49%.

→ Strong, statistically significant predictor of poorer survival ($p < 0.001$).

Intercept (0.029)

→ Not directly interpretable as a hazard ratio (it sets the baseline rate), so usually ignored.

Take-home interpretation

Treatment is protective (HR < 1).

Sex (male), age, and especially biomarker are risk factors (HR > 1).

Biomarker has the strongest impact: nearly 1.5× increase in hazard per unit.

Exponential model via flexsurvreg (rate/PH form)

Coefficients here are on the log-rate (log-hazard) scale; exp(coef) gives hazard ratios directly.

```
fit_ph <- flexsurvreg(Surv(time, status) ~ treatment + sex + age + biomarker,
                      data = dat, dist = "exponential")
fit_ph
```

```
## Call:
## flexsurvreg(formula = Surv(time, status) ~ treatment + sex +
##     age + biomarker, data = dat, dist = "exponential")
##
## Estimates:
##             data mean  est        L95%       U95%       se          exp(est)
## rate               NA   0.029447   0.011329   0.076538   0.014351
NA
## treatment   0.520000  -0.313738  -0.647563   0.020088   0.170322
0.730711
## sex         0.506667   0.202053  -0.128801   0.532908   0.168806
1.223913
## age        54.245000   0.015900  -0.000466   0.032266   0.008350
1.016027
## biomarker   0.172763   0.395474   0.215043   0.575905   0.092058
1.485088
##             L95%       U95%
## rate               NA         NA
## treatment   0.523320   1.020291
## sex         0.879149   1.703879
## age         0.999535   1.032792
## biomarker   1.239915   1.778740
##
## N = 300,  Events: 141,  Censored: 159
## Total time at risk: 1976.827
## Log-likelihood = -499.4238, df = 5
## AIC = 1008.848
```

```
exp(coef(fit_ph))  # hazard ratios
```

```
##       rate  treatment        sex        age  biomarker
## 0.02944704 0.73071081 1.22391333 1.01602736 1.48508818
```

Baseline

rate = 0.029 (95% CI: 0.011 – 0.077)

→ This is the baseline hazard rate per unit time for a “reference” patient (when covariates = 0).

It sets the overall time scale but is usually less interpretable than HRs.

Treatment (HR 0.73)

Patients on treatment have about a 27% lower hazard of death compared to controls. Borderline significant.

Sex (male, HR 1.22)

Males have ~22% higher hazard compared to females, but this is not statistically significant.

Age (HR 1.02 per year)

Each additional year increases hazard by about 2%. Over 10 years, hazard increases by ~22% (1.02^10 ≈ 1.22). Borderline significant.

Biomarker (HR 1.49)

Each +1 unit of the biomarker increases hazard by 49%. This is the strongest and most significant predictor in your model.

Model fit

Events = 141, Censored = 159 (about half censored, which is fine).

Log-likelihood = -499.4, AIC = 1008.8 → this quantifies model fit; useful mainly for comparing models (e.g., exponential vs Weibull).

Summary (practical interpretation)

The treatment is protective (HR < 1), though borderline significant.

Biomarker is the key risk factor, significantly raising hazard (HR ~1.5).

Older age is linked to higher hazard (borderline).

Sex (male) shows a higher hazard but is not statistically significant.

In short: Treatment helps, but biomarker strongly drives risk, with age also playing a modest role.

Predicted survival curves by group (flexsurv)

Example: treatment=0 vs 1 at median covariate values

```
newdat <- data.frame(
  treatment = c(0,1),
  sex = median(dat$sex),
  age = median(dat$age),
  biomarker = median(dat$biomarker)
)
```

Survival at specific times

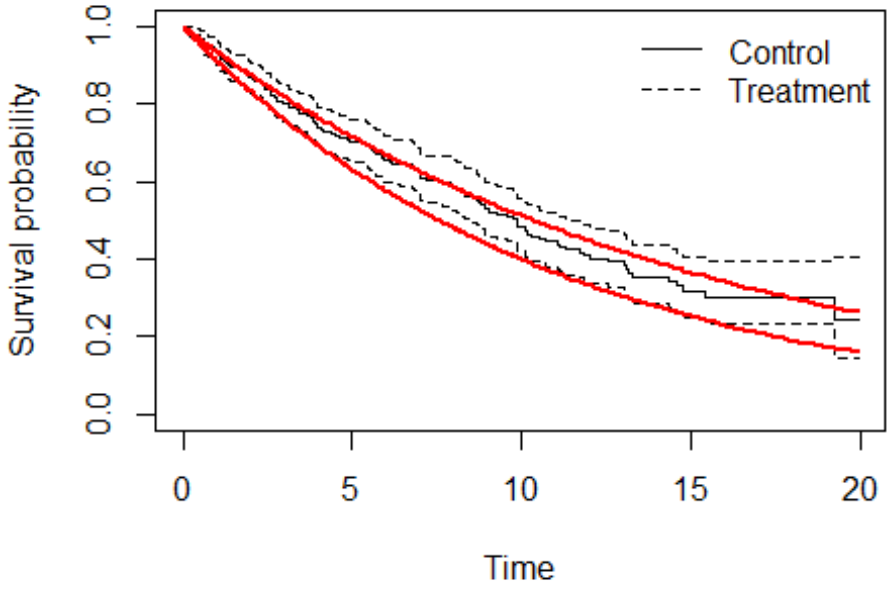
```
predict(fit_ph, newdata = newdat, type = "survival", times = c(2,5,10))
```

```
## # A tibble: 2 × 1
##   .pred
##   <list>
## 1 <tibble [3 × 2]>
## 2 <tibble [3 × 2]>
```

simple plot of model-based survival curves

(Base R quickplot)

```
plot(fit_ph, type = "survival", newdata = newdat, ci = FALSE,
     xlab = "Time", ylab = "Survival probability")
legend("topright", legend = c("Control","Treatment"), lty = 1:2, bty = "n")
```


Survival probability
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
Time
Control
Treatment