LOG LOGISTIC SURVIVAL MODEL - AFT

Enock Bereka

2025-08-25

The log-logistic survival model is a parametric survival model where survival times are assumed to follow a log-logistic distribution. It is commonly used in survival analysis when the hazard rate is non-monotonic (it first increases, reaches a peak, and then decreases).

Load required packages

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

Load the data

```
d <- read_csv("C:/Users/ADMIN/Desktop/Data
Science/Datasets/survival/loglogistic_simulated_survival.csv")
```

Recode categorical variables

```
d$sex <- ifelse(d$sex == 1, "Male", "Female")
d$sex <- factor(d$sex)
d$treatment <- ifelse(d$treatment == 1, "drug", "placebo")
d$treatment <- factor(d$treatment)
```

Quick check

```
glimpse(d)

## Rows: 600
## Columns: 6
## $ time      <dbl> 4.5587, 1.5165, 1.7022, 6.7065, 27.3270, 20.3171,
6.6724, 0.…
## $ event     <dbl> 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0, 0, 0, 1, 1,
1, …
## $ age       <dbl> 65.0, 58.6, 66.5, 75.2, 57.7, 57.7, 75.8, 67.7, 55.3,
65.4, …
## $ sex       <fct> Female, Male, Male, Male, Male, Male, Female, Female,
Female…
## $ treatment <fct> placebo, drug, placebo, drug, placebo, placebo, drug,
placeb…
## $ biomarker <dbl> 0.756, 0.384, 1.361, 0.726, 0.496, 0.178, 0.984, 2.896,
2.70…

with(d, table(event))
```

```
## event
##   0   1
## 332 268
```

——————- AFT MODELS ——————-

AFT via flexsurv (log-logistic)

```
aft_llogis <- flexsurvreg(Surv(time, event) ~ age + sex + treatment +
biomarker,
                          data = d, dist = "llogis")
print(aft_llogis)

## Call:
## flexsurvreg(formula = Surv(time, event) ~ age + sex + treatment +
##     biomarker, data = d, dist = "llogis")
##
## Estimates:
##                    data mean  est       L95%      U95%      se
exp(est)
## shape                    NA    1.52611   1.38555   1.68092   0.07523
NA
## scale                    NA   25.65934  12.47679  52.77013   9.43964
NA
## age                59.86467   -0.01563  -0.02678  -0.00448   0.00569
0.98450
## sexMale             0.50667    0.15691  -0.06173   0.37556   0.11156
1.16989
## treatmentplacebo    0.48833    0.32868   0.10956   0.54780   0.11180
1.38913
## biomarker           1.19133    0.22254   0.07882   0.36626   0.07333
1.24924
##                    L95%      U95%
## shape                    NA        NA
## scale                    NA        NA
## age                 0.97358   0.99553
## sexMale             0.94014   1.45580
## treatmentplacebo    1.11579   1.72944
## biomarker           1.08201   1.44233
##
## N = 600,  Events: 268,  Censored: 332
## Total time at risk: 6998.104
## Log-likelihood = -1119.203, df = 6
## AIC = 2250.406
```

Model Setup

Model: Log-logistic Accelerated Failure Time (AFT) model

Outcome: Survival time (Surv(time, event))

Predictors: age, sex, treatment, biomarker

Distribution Parameters

Shape = 1.526 (95% CI: 1.39 – 1.68)

Controls how “peaked” or “spread out” the survival curve is.

1 means the hazard initially increases with time, peaks, then decreases.

Scale = 25.66 (95% CI: 12.48 – 52.77)

This is a baseline time parameter. Higher scale shifts survival curve to longer survival times.

Covariates (AFT interpretation)

In AFT models, coefficients represent the log acceleration factor.

exp(est) = acceleration factor (time ratio).

1: longer survival (protective).

<1: shorter survival (risk factor).

Age: The estimated coefficient for age is -0.0156, with exp(β) = 0.985. This means that for each 1-year increase in age, survival time is expected to be about 1.5% shorter. The effect is small but statistically significant.

Sex (Male vs Female): The coefficient for males is 0.157, with exp(β) = 1.17. This suggests that males have about 17% longer survival compared to females. However, since the 95% confidence interval (0.94–1.46) includes 1, this effect is not statistically significant.

Treatment (Placebo vs Drug): The treatment coefficient is 0.329, with exp(β) = 1.389. Patients receiving placebo appear to have about 39% longer survival compared to those receiving the drug. The confidence interval (1.12–1.73) indicates that this effect is statistically significant.

Biomarker: The coefficient for the biomarker is 0.223, with exp(β) = 1.249. Each unit increase in the biomarker is associated with about 25% longer survival. The confidence interval (1.08–1.44) shows this is a statistically significant effect.

Model Fit

N = 600 (268 deaths, 332 censored).

Log-likelihood = -1119.2.

AIC = 2250.4 → useful for comparing with other models (e.g., Weibull, log-normal).

Key Takeaways

Older age → shorter survival.

Treatment effect is surprising: placebo patients lived longer than treated ones (maybe treatment is harmful, or sicker patients got treated).

Biomarker → strong positive predictor of survival.

Sex → not statistically significant.

Log-logistic fit: hazard is non-monotonic (rises, then falls), so it fits diseases/events where risk peaks then declines.

AFT via survreg (parametric AFT; different parameterization)

```
aft_llogis_sr <- survreg(Surv(time, event) ~ age + sex + treatment +
biomarker,
                         data = d, dist = "loglogistic")
summary(aft_llogis_sr)

##
## Call:
## survreg(formula = Surv(time, event) ~ age + sex + treatment +
##     biomarker, data = d, dist = "loglogistic")
```

```
##                       Value Std. Error     z      p
## (Intercept)       3.24491    0.36771  8.82 <2e-16
## age              -0.01563    0.00569 -2.75 0.0060
## sexMale           0.15691    0.11156  1.41 0.1595
## treatmentplacebo  0.32868    0.11180  2.94 0.0033
## biomarker         0.22254    0.07333  3.03 0.0024
## Log(scale)       -0.42272    0.04929 -8.58 <2e-16
##
## Scale= 0.655
##
## Log logistic distribution
## Loglik(model)= -1119.2   Loglik(intercept only)= -1132.5
##  Chisq= 26.5 on 4 degrees of freedom, p= 2.5e-05
## Number of Newton-Raphson Iterations: 4
## n= 600
```

Age ($\beta = -0.0156$, $p = 0.006$)

For every 1-year increase in age, survival time is multiplied by $\exp(-0.0156) \approx 0.985$.

This means older patients have ~1.5% shorter survival time per year of age. Effect is statistically significant.

Sex (Male vs Female; $\beta = 0.157$, $p = 0.16$)

Males have a time ratio of $\exp(0.157) \approx 1.17$.

This suggests that males live ~17% longer than females, but the result is not statistically significant ($p > 0.05$).

Treatment (Placebo vs Drug; $\beta = 0.329$, $p = 0.0033$)

Patients on placebo have a time ratio of $\exp(0.329) \approx 1.39$.

This means placebo patients survive ~39% longer than those on the drug, and this effect is statistically significant.

Biomarker ($\beta = 0.223$, $p = 0.0024$)

Each unit increase in biomarker is associated with a time ratio of $\exp(0.223) \approx 1.25$.

So, higher biomarker levels correspond to ~25% longer survival time, and the effect is significant.

Scale parameter (Scale = 0.655, log(scale) = -0.423, $p < 2e-16$)

This describes variability in survival times. A scale < 1 suggests that the hazard function rises quickly and then falls off (common with log-logistic).

Overall model fit

Likelihood ratio test: $\chi^2(4) = 26.5$, $p < 0.0001$ → The covariates collectively improve model fit compared to intercept-only.

Sample size: $n = 600$, so findings are reasonably powered.

In short:

Older age shortens survival.

Males tend to live longer, but not significantly.

Placebo group survives longer than drug group (significant).

Higher biomarker values predict longer survival (significant).

————————- Prediction ————————-

Create a representative dataset for predictions

```
newdat_aft <- d %>%
  group_by(treatment) %>%
  summarize(
    age = mean(age),
    biomarker = median(biomarker),
    sex = names(sort(table(sex), decreasing = TRUE))[1],
    .groups = "drop"
  ) %>%
  mutate(sex = factor(sex, levels = levels(d$sex)))

print(newdat_aft)

## # A tibble: 2 × 4
##   treatment   age biomarker sex
##   <fct>     <dbl>     <dbl> <fct>
## 1 drug       59.7     0.994 Male
## 2 placebo    60.0     0.989 Female
```

Predict with flexsurv: survival estimates over a grid of times

```
tgrid <- seq(0, quantile(d$time, 0.99), length.out = 200)

aft_surv_list <- lapply(split(newdat_aft, newdat_aft$treatment), function(nd)
{
  s <- summary(aft_llogis, newdata = nd, type = "survival", t = tgrid)
  data.frame(time = tgrid, surv = s[[1]]$est, treatment = nd$treatment)
})

aft_surv_df <- do.call(rbind, aft_surv_list)
```

————————- Plot survival curves ————————-

```
ggplot(aft_surv_df, aes(x = time, y = surv, linetype = treatment)) +
  geom_step(size = 1.2) +
  labs(title = "AFT (Log-logistic) — Predicted Survival by Treatment",
       x = "Time", y = "Survival Probability") +
```

```r
  theme_minimal(base_size = 14) +
  theme(legend.title = element_blank())
```