LOG NORMAL SURVIVAL MODEL - AFT

Enock Bereka

2025-08-26

Load required packages

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

Load the data

```r
dat <- read_csv("C:/Users/ADMIN/Desktop/Data
Science/Datasets/survival/sim_lognormal_survival.csv")
```

Quick check

```r
glimpse(dat)
```

```
## Rows: 500
## Columns: 8
## $ id          <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16,
17,…
## $ time        <dbl> 4.3526591, 2.5375313, 4.5170621, 3.8048794, 1.6090646,
1.3…
## $ status      <dbl> 1, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 1, 1, 0, 0, 0, 1, 0,
0, 1…
## $ age         <dbl> 59.1, 67.3, 45.6, 53.4, 59.0, 81.5, 73.9, 57.9, 67.6,
62.1…
## $ treatment   <chr> "control", "treated", "control", "control", "control",
"co…
## $ biomarker   <dbl> 0.317, -0.710, -0.559, -1.020, 0.451, -0.555, 0.655,
0.705…
## $ T_true      <dbl> 4.353, 4.798, 4.517, 4.026, 2.306, 2.000, 6.812,
3.363, 4.…
## $ censor_time <dbl> 6.102, 2.538, 5.798, 3.805, 1.609, 1.320, 1.253,
2.066, 4.…
```

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

```
## status
##   0   1
## 250 250
```

Fit log-normal Accelerated Failure Time (AFT) model

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

Model summary (coefficients are on log(time) scale)

```
summary(fit_aft)

##
## Call:
## survreg(formula = Surv(time, status) ~ treatment + age + biomarker,
##     data = dat, dist = "lognormal")
##                     Value Std. Error      z       p
## (Intercept)       2.47586    0.20399  12.14 < 2e-16
## treatmenttreated  0.49320    0.06566   7.51 5.8e-14
## age              -0.01907    0.00334  -5.70 1.2e-08
## biomarker        -0.30202    0.03220  -9.38 < 2e-16
## Log(scale)       -0.51823    0.04518 -11.47 < 2e-16
##
## Scale= 0.596
##
## Log Normal distribution
## Loglik(model)= -612.5   Loglik(intercept only)= -687.5
##  Chisq= 150 on 3 degrees of freedom, p= 2.6e-32
## Number of Newton-Raphson Iterations: 5
## n= 500
```

Extract time ratios (exp(coefficients))

Regression coefficients (only the covariates)

```
est <- coef(fit_aft)
```

Standard errors for just those coefficients

```
se <- sqrt(diag(vcov(fit_aft)))[names(est)]
```

Compute time ratios + 95% CI

```
time_ratio <- exp(est)
ci_low <- exp(est - 1.96 * se)
ci_high <- exp(est + 1.96 * se)
```

Put into a nice table

```
results <- data.frame(
  term = names(est),
  coef = est,
  se = se,
  time_ratio = time_ratio,
  CI_lower = ci_low,
  CI_upper = ci_high,
  row.names = NULL
)
```

```
library(flextable)
flextable(results) %>% theme_box()
```

term	coef	se	time_ratio	CI_lower	CI_upper
(Intercept)	2.47585876	0.20398961 4	11.8919150	7.9728269	17.7374532
treatmentt reated	0.49319715	0.06565563 1	1.6375433	1.4398114	1.8624302
age	- 0.01906664	0.00334382 8	0.9811140	0.9747049	0.9875652
biomarker	- 0.30201580	0.03220206 9	0.7393264	0.6941051	0.7874938

Coefficients

(Intercept) = 2.476 (p < 0.001)

Baseline log(survival time) when treatment = control, age = 0, biomarker = 0.

(Not usually directly interpretable since age=0 isn't realistic, but necessary for model structure.)

treatmenttreated = 0.493 (p < 0.001)

On the log-time scale.

Exponentiate: exp(0.493) ≈ 1.64.

Treated patients live ~64% longer than controls, all else equal.

age = -0.019 (p < 0.001)

Each additional year of age reduces log-time by 0.019.

Exponentiate: exp(-0.019) ≈ 0.981.

Every extra year of age shortens survival time by about 2%.

biomarker = -0.302 (p < 0.001)

Each 1-unit increase in the biomarker decreases log-time by 0.302.

Exponentiate: exp(-0.302) ≈ 0.74.

Higher biomarker values are associated with 26% shorter survival times.

Scale parameter

Log(scale) = -0.518 → Scale = 0.596

This is the standard deviation of the log-survival times (σ in the log-normal).

Smaller values = less variability around the regression prediction.

Model fit statistics

Loglik(model) = -612.5 vs intercept-only = -687.5

Likelihood Ratio Test: χ^2 = 150, df = 3, p ≈ 2.6e-32

Strong evidence that covariates (treatment, age, biomarker) significantly improve prediction of survival.

n = 500 observations were used.

Plain language summary

The treatment effect is large and highly significant: treated patients live about 1.6× longer than controls.

Older age is associated with shorter survival (about 2% reduction per year).

High biomarker values are strongly predictive of poorer survival (about 26% reduction per unit increase).

The model explains survival well (p-value < 0.0001).

predicted median survival for a patient

```r
new_patient <- data.frame(treatment = "treated", age = 60, biomarker = 0)
predicted_logmed <- predict(fit_aft, newdata = new_patient, type =
"quantile", p = 0.5)
predicted_median <- exp(predicted_logmed)

cat("Predicted median survival (treated, age=60, biomarker=0):",
    round(predicted_median, 2), "\n")

## Predicted median survival (treated, age=60, biomarker=0): 494.3
```