WEIBULL MODEL - AFT

Enock Bereka

2025-08-25

A Weibull model is a type of parametric survival analysis model used to study the time until an event occurs (such as death, relapse, machine failure, etc.). It assumes that survival times follow a Weibull distribution, which is very flexible because it can model increasing, decreasing, or constant hazard rates depending on the shape parameter.

Load required packaeges

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

Load the data

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

Quick check

```
glimpse(dat)
```

```
## Rows: 300
## Columns: 7
## $ id        <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16,
17, 1…
## $ time      <dbl> 25.61, 4.87, 25.34, 26.23, 12.04, 2.23, 7.52, 3.49,
17.36, 8…
## $ status    <int> 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 1, 0, 0, 1, 0, 0, 1, 1, 0,
1, …
## $ treatment <int> 1, 0, 0, 1, 1, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1,
1, …
## $ sex       <int> 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 0, 0, 1, 1, 1, 1,
1, …
## $ age       <dbl> 72.7, 72.4, 85.0, 53.1, 42.0, 57.2, 50.1, 56.9, 47.4,
66.8, …
## $ biomarker <dbl> -0.69, -0.06, -1.16, 0.13, -1.36, 0.77, -1.06, 1.32, -
0.01, …
```

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

```
## status
##   0   1
## 164 136
```

Weibull AFT model with survreg

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

##
## Call:
## survreg(formula = Surv(time, status) ~ treatment + sex + age +
##     biomarker, data = dat, dist = "weibull")
##                Value Std. Error     z       p
## (Intercept)  3.22573    0.33877  9.52 < 2e-16
## treatment    0.77420    0.13653  5.67 1.4e-08
## sex         -0.31300    0.12347 -2.53   0.011
## age         -0.00438    0.00515 -0.85   0.396
## biomarker   -0.51019    0.07610 -6.70 2.0e-11
## Log(scale)  -0.33593    0.07062 -4.76 2.0e-06
##
## Scale= 0.715
##
## Weibull distribution
## Loglik(model)= -544.5   Loglik(intercept only)= -587.2
##  Chisq= 85.34 on 4 degrees of freedom, p= 1.3e-17
## Number of Newton-Raphson Iterations: 5
## n= 300
```

Model type

Distribution: Weibull.

This is an AFT model, meaning coefficients describe how covariates accelerate or decelerate survival time.

Positive coefficient → longer survival (protective).

Negative coefficient → shorter survival (risk factor).

Coefficients

Intercept (3.23): baseline log survival time (not directly meaningful).

Treatment (0.77, $p < 0.001$)

→ Strong positive effect.

→ exp(0.774) ≈ 2.17 → treatment patients live about 2.2× longer (median survival) than controls.

Sex (-0.31, $p = 0.011$)

→ Males have shorter survival.

→ exp(-0.313) ≈ 0.73 → survival time is 27% shorter compared to females.

Age (-0.004, $p = 0.396$)

→ Not statistically significant.

→ Each year of age reduces survival time by only ~0.4%.

Biomarker (-0.51, $p < 0.001$)

→ Strong negative effect.

→ exp(-0.51) ≈ 0.60 → higher biomarker values reduce survival by 40% per unit.

Scale Parameter

Interpretation:

k > 1 (here 1.4) → hazard increases with time (aging or disease progression risk rises).

If k = 1, it reduces to the exponential model (constant hazard).

So your data suggest risk of failure grows over time.

Model fit

Log-likelihood = -544.5 vs intercept-only = -587.2

Likelihood ratio $\chi^2 = 85.3$, $p < 0.0001$

→ Covariates significantly improve fit.

Practical summary

Treatment greatly prolongs survival (~2.2× longer).

Males have worse survival (~27% shorter).

Biomarker is the strongest risk factor (40% shorter survival per unit increase).

Age shows no significant effect in this model.

Shape parameter (k ≈ 1.4) tells us hazard increases over time → Weibull fits better than exponential here.

Weibull PH model with flexsurv

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

```
## Call:
## flexsurvreg(formula = Surv(time, status) ~ treatment + sex +
##     age + biomarker, data = dat, dist = "weibull")
##
## Estimates:
##             data mean  est       L95%      U95%      se        exp(est)
## shape             NA    1.39925   1.21837   1.60698   0.09882        NA
## scale             NA   25.17196  12.95842  48.89698   8.52761        NA
## treatment    0.50667    0.77420   0.50661   1.04178   0.13653   2.16885
## sex          0.48333   -0.31300  -0.55501  -0.07099   0.12347   0.73125
```

```
## age        60.22800   -0.00438  -0.01448   0.00573   0.00515   0.99563
## biomarker  -0.09843   -0.51019  -0.65934  -0.36104   0.07610   0.60038
##            L95%      U95%
## shape            NA        NA
## scale            NA        NA
## treatment   1.65965   2.83426
## sex         0.57407   0.93147
## age         0.98562   1.00574
## biomarker   0.51719   0.69695
##
## N = 300,  Events: 136,  Censored: 164
## Total time at risk: 3844.67
## Log-likelihood = -544.5485, df = 6
## AIC = 1101.097
```

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

```
##      shape      scale  treatment        sex        age  biomarker
##  1.3992459 25.1719640  2.1688467  0.7312499  0.9956316  0.6003790
```

Covariate Effects (exp(est) = Hazard Ratio)

These are interpreted like hazard ratios in Cox regression:

Treatment

HR = 2.17 (95% CI: 1.66 – 2.83)

Patients on treatment have more than twice the hazard (risk of event) compared to the control group.

Suggests treatment is harmful in this simulated dataset.

Sex

HR = 0.73 (95% CI: 0.57 – 0.93)

Being in the coded sex group (likely female if male=0, female=1) reduces hazard by ~27%.

That group has better survival.

Age

HR = 0.996 (95% CI: 0.986 – 1.006)

Very close to 1, not statistically significant.

No strong effect of age on survival in this dataset.

Biomarker

HR = 0.60 (95% CI: 0.52 – 0.70)

Higher biomarker levels are associated with a 40% lower hazard.

Suggests biomarker is protective.

Interpretation in Plain Words

The hazard of dying increases over time (Weibull shape > 1).

Treatment increases risk, possibly harmful.

One sex (coded as “1”) survives longer than the other.

Biomarker protects against death.

Age doesn’t play a big role here (not significant).

Survival predictions

```
newdat <- data.frame(
  treatment = c(0,1),
  sex = 0,
  age = 60,
  biomarker = 0
)

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")
```