

# Weibull AFT Regression Functions in R

Sarah R. Haile\*

October 21, 2024

Weibull accelerated failure time regression can be performed in R using the `survreg` function. The results are not, however, presented in a form in which the Weibull distribution is usually given. Accelerated failure time models are usually given by

$$\log T = Y = \mu + \boldsymbol{\alpha}^T \mathbf{z} + \sigma W,$$

where  $\mathbf{z}$  are set of covariates, and  $W$  has the extreme value distribution. Given transformations

$$\begin{aligned}\gamma &= 1/\sigma, \\ \lambda &= \exp(-\mu/\sigma), \\ \boldsymbol{\beta} &= -\boldsymbol{\alpha}/\sigma,\end{aligned}$$

we have a Weibull model with baseline hazard of

$$h(x|\mathbf{z}) = (\gamma \lambda t^{\gamma-1}) \exp(\boldsymbol{\beta}^T \mathbf{z}).$$

Further, the `survreg` function generally gives  $\log \sigma$ , rather than  $\sigma$  as output. The function `WeibullReg` (along with `ConvertWeibull`) solves this problem. Hazard ratios ( $\exp(\boldsymbol{\beta}_i)$ ) are additionally produced.

The function also produces the “event time ratio” (ETR,  $\exp(-\boldsymbol{\beta}_i/\gamma) = \exp \boldsymbol{\alpha}_i$ ), as discussed in [1]. This ratio quantifies the relative difference in time it takes to achieve the  $p$ th percentile between two levels of a covariate. The  $p$ th percentile of the (covariate-adjusted) Weibull distribution occurs at

$$t_p = \left[ \frac{-\log p}{\lambda e^{\boldsymbol{\beta}^T \mathbf{z}}} \right]^{1/\gamma}.$$

Then the ratio of times for a covariate with value  $z_1$  versus values  $z_0$ , with parameter estimate  $\beta$ , can then be computed as:

$$\begin{aligned}\frac{t_B}{t_A} &= \left[ \frac{-\log p}{\lambda e^{\boldsymbol{\beta} z_1}} \right]^{1/\gamma} \left[ \frac{\lambda e^{\boldsymbol{\beta} z_0}}{-\log p} \right]^{1/\gamma} \\ &= \exp \left\{ \frac{\boldsymbol{\beta}(z_0 - z_1)}{\gamma} \right\}.\end{aligned}$$

Thus, if we are comparing treatment B to treatment A, where the parameter estimate for treatment B is  $\boldsymbol{\beta}_{\text{trt}}$ , then the ETR is  $\exp\{-\boldsymbol{\beta}_{\text{trt}}/\gamma\}$ .

---

\*Epidemiology, Biostatistics and Prevention Institute (EBPI), University of Zurich (sarah.haile@uzh.ch)

For example if the ETR for treatments A vs B is 1.2, then the amount of time it takes for  $p$  percent of patients with treatment A to have the event is predicted to be about 20% longer than it takes for the same percentage of patients with treatment B to experience an event. (That is, treatment B is worse.) For this reason, the ETR can also be called an “acceleration factor.”

Additionally, a function `WeibullDiag` has been provided to check the adequacy of the Weibull Model.

## 1 WeibullReg

The `WeibullReg` function performs Weibull AFT regression on survival data, returning a list which contains:

**formula** the regression formula,

**coef** the coefficient table,

**HR** a table with the hazard rates (with confidence intervals) for each of the covariates,

**ETR** a table with the Event Time Ratios (with confidence intervals) for each of the covariates, and

**summary** the summary table from the original `survreg` model.

Such tables can also be produced using the `streg` function in `stata` with the following options: 1) the `nohr` option gives `coef`, 2) without any options gives `HR`, 3) the `tr` option gives `ETR`, and 4) the `time` option produces `summary`, the original output from `survreg`. While `proc lifereg` in `SAS` can also perform parametric regression for survival data, its output must also be transformed.

The following example reproduces Tables 12.1 and 12.2 from [2], on the `larynx` data set.

```
> library(survival)
> data(larynx)

> WeibullReg(Surv(time, death) ~ factor(stage) + age, data=larynx)

$formula
Surv(time, death) ~ factor(stage) + age

$coef
              Estimate      SE
lambda      0.01853664 0.01898690
gamma       1.13014371 0.13844846
factor(stage)2 0.16692694 0.46112943
factor(stage)3 0.66289534 0.35550887
factor(stage)4 1.74502788 0.41476410
age          0.01973646 0.01424135

$HR
              HR      LB      UB
factor(stage)2 1.181668 0.4786096 2.917491
```

```

factor(stage)3 1.940402 0.9666786 3.894946
factor(stage)4 5.726061 2.5398504 12.909334
age            1.019933 0.9918573 1.048802

```

\$ETR

```

                ETR          LB          UB
factor(stage)2 0.8626863 0.3880879 1.917678
factor(stage)3 0.5562383 0.2971113 1.041364
factor(stage)4 0.2135090 0.1047619 0.435140
age            0.9826879 0.9583820 1.007610

```

\$summary

Call:

```
survival::survreg(formula = formula, data = data, dist = "weibull")
```

```

                Value Std. Error      z      p
(Intercept)    3.5288    0.9041  3.90 9.5e-05
factor(stage)2 -0.1477    0.4076 -0.36  0.717
factor(stage)3 -0.5866    0.3199 -1.83  0.067
factor(stage)4 -1.5441    0.3633 -4.25 2.1e-05
age             -0.0175    0.0128 -1.37  0.172
Log(scale)     -0.1223    0.1225 -1.00  0.318

```

Scale= 0.885

Weibull distribution

```

Loglik(model)= -141.4  Loglik(intercept only)= -151.1
      Chisq= 19.37 on 4 degrees of freedom, p= 0.00066
Number of Newton-Raphson Iterations: 5
n= 90

```

The hazard rates produced with the Weibull regression model are similar to what is obtained with Cox proportional hazards regression:

```

> cph <- coxph(Surv(time, death) ~ factor(stage) + age, data=larynx)
> summary(cph)$conf.int

```

```

                exp(coef) exp(-coef) lower .95 upper .95
factor(stage)2  1.150320  0.8693233 0.4646755  2.847656
factor(stage)3  1.901003  0.5260381 0.9459343  3.820364
factor(stage)4  5.506778  0.1815944 2.4085976 12.590147
age             1.019213  0.9811488 0.9911247  1.048098

```

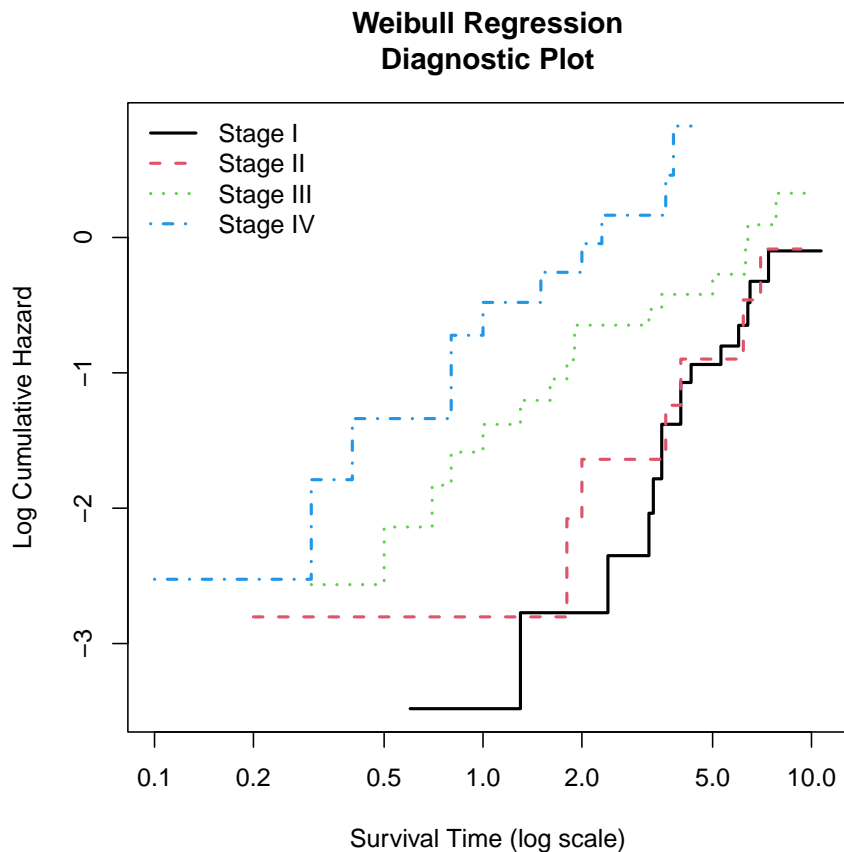
Most of the work of the function is actually performed by `ConvertWeibull`. These functions require the `survival` package in R. Formulas for the variance estimates come from [2, Equations 12.2.13-18, with some modifications since R gives  $\log \sigma$ ].

## 2 WeibullDiag

The `WeibullDiag` function produces a diagnostic plot for Weibull AFT regression, similar to what is in [2, Figure 12.2]. It plots  $\log$  Time versus the  $\log$  of the estimated cumulative

hazard estimate. If the Weibull model has adequate fit, then the plots for each of the covariates should be roughly linear and parallel. This function requires the `survfit` object to contain `strata`, else an error is produced. The `WeibullDiag` function requires the `survival` package.

```
> WeibullDiag(Surv(time, death) ~ factor(stage), data = larynx,  
+             labels=c("Stage I", "Stage II", "Stage III", "Stage IV"))
```



## References

- [1] K. J. Carroll. On the use and utility of the Weibull model in the analysis of survival data. *Controlled Clinical Trials*, 24(6):682 – 701, 2003.
- [2] J. P. Klein and M. L. Moeschberger. *Survival Analysis: techniques for censored and truncated data*. Springer Verlag, 2003.