

# Weibull accelerated failure time regression Regression

Sarah R. Haile\*

November 3, 2023

Weibull accelerated failure time (AFT) 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

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

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

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

we have a Weibull model with baseline hazard of

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

Further, the **survreg** function generally gives  $\log \sigma$ , rather than  $\sigma$  as output. The function **weibullReg** (along with **survreg2weibull**) solves this problem. Hazard ratios ( $\exp(\boldsymbol{\beta}_i)$ ) are additionally produced.

The function also produces the “event time ratio” (also known as “acceleration factor”,  $\exp(-\boldsymbol{\beta}_i/\alpha) = \exp \boldsymbol{\gamma}_i$ ), as discussed in Carroll (2003). 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

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

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

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

---

\*University of Zurich, Institute for Social and Preventive Medicine, Biostatistics Unit  
(sarah.haile@ifspm.uzh.ch)

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

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 Klein and Moeschberger (2003), on the `larynx` data set.

```
> library(biostatUZH)
> data(larynx)
> larynx$stage2 <- as.numeric(larynx$stage == 2)
> larynx$stage3 <- as.numeric(larynx$stage == 3)
> larynx$stage4 <- as.numeric(larynx$stage == 4)

> weibullReg(Surv(time, death) ~ stage2 + stage3 + stage4 + age, data=larynx)

$formula
Surv(time, death) ~ stage2 + stage3 + stage4 + age

$coef
      Estimate      SE
lambda 0.01853664 0.01898690
alpha  1.13014371 0.13844846
```

```
stage2 0.16692694 0.46112943
stage3 0.66289534 0.35550887
stage4 1.74502788 0.41476410
age     0.01973646 0.01424135
```

\$HR

	HR	LB	UB
stage2	1.181668	0.4786096	2.917491
stage3	1.940402	0.9666786	3.894946
stage4	5.726061	2.5398504	12.909334
age	1.019933	0.9918573	1.048802

\$ETR

	ETR	LB	UB
stage2	0.8626863	0.3880879	1.917678
stage3	0.5562383	0.2971113	1.041364
stage4	0.2135090	0.1047619	0.435140
age	0.9826879	0.9583820	1.007610

\$summary

Call:

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

	Value	Std. Error	z	p
(Intercept)	3.5288	0.9041	3.90	9.5e-05
stage2	-0.1477	0.4076	-0.36	0.717
stage3	-0.5866	0.3199	-1.83	0.067
stage4	-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:

```
> summary(coxph(Surv(time, death) ~ stage2 + stage3 + stage4 + age,
+               data=larynx))$conf.int
```

	exp(coef)	exp(-coef)	lower .95	upper .95
stage2	1.150320	0.8693233	0.4646755	2.847656
stage3	1.901003	0.5260381	0.9459343	3.820364

```
stage4  5.506778  0.1815944  2.4085976  12.590147
age      1.019213  0.9811488  0.9911247   1.048098
```

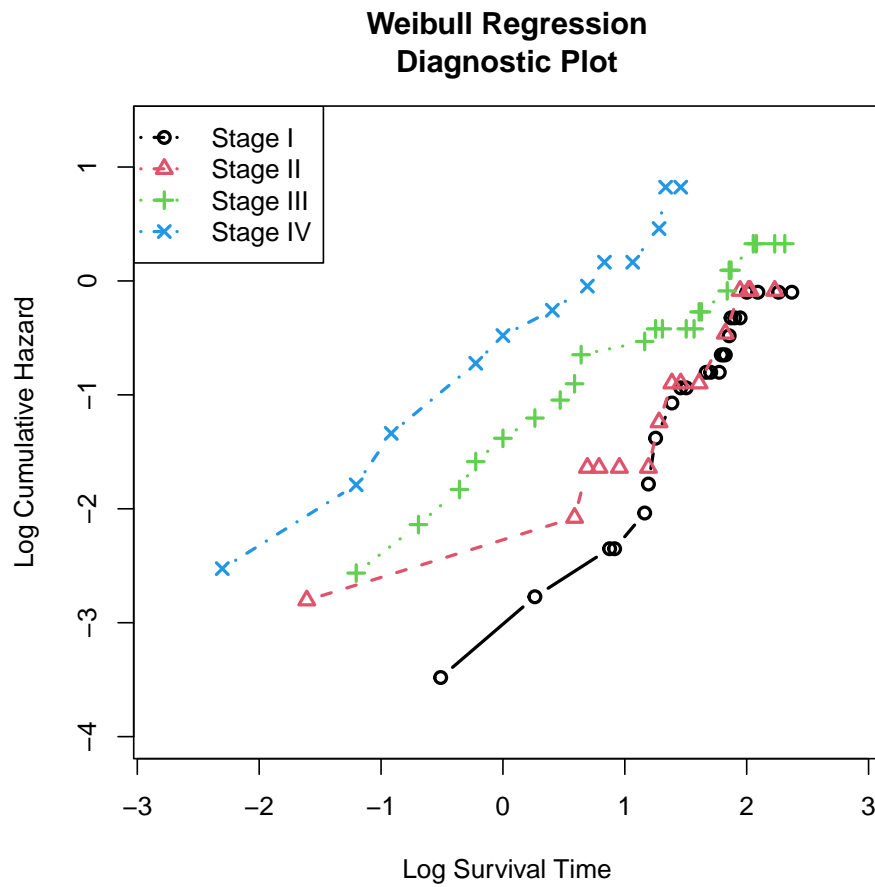
Most of the work of the function is actually performed by `survreg2weibull`. These functions require the `survival` package in R. Formulas for the variance estimates come from Klein and Moeschberger (2003) (Equations 12.2.13-18, with some modifications since R gives  $\log \sigma$ ).

## 2 weibullDiag

The `weibullDiag` function produces a diagnostic plot for Weibull AFT regression, similar to what is present Klein and Moeschberger (2003), Figure 12.2. It plots  $\log \text{Time}$  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, at the moment, only works for one nominal covariate (or a set covariates that defines one factor). For example, if disease stage can be divided into 4 categories, one covariate can be used with levels 1:4, or alternately, 3 binary covariates. (The labels for the two types of covariates may vary, however.) It returns a table with the data used to create the plot.

This function requires both `survival` and `prodlim`.

```
> fm <- Surv(time, death) ~ stage
> weibullDiag(formula = fm, data = larynx,
+             labels = c("Stage I", "Stage II", "Stage III", "Stage IV"))
```



## References

- CARROLL, K. (2003). On the use and utility of the weibull model in the analysis of survival data. *Controlled Clinical Trials*, **24** 682–701.
- KLEIN, J. and MOESCHBERGER, M. (2003). *Survival analysis: techniques for censored and truncated data*. Springer.