Weibull AFT Regression Functions in R

Sarah R. Haile∗

October 8, 2015

Weibull accelerated failure time regression can be performed in R using the \texttt{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
\[ \log T = Y = \mu + \alpha^T z + \sigma W, \]
where \( z \) are set of covariates, and \( W \) has the extreme value distribution. Given transformations
\[ \gamma = 1/\sigma, \]
\[ \lambda = \exp(-\mu/\sigma), \]
\[ \beta = -\alpha/\sigma, \]
we have a Weibull model with baseline hazard of
\[ h(x|z) = (\gamma \lambda t^{\gamma-1}) \exp(\beta^T z). \]

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

The function also produces the “event time ratio” (ETR, \( \exp(-\beta_i/\gamma) = \exp(\alpha_i) \)), as discussed in ?. 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
\[ t_p = \left[ -\log p / \lambda e^{\beta^T z} \right]^{1/\gamma}. \]

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:
\[ \frac{t_B}{t_A} = \left[ \frac{-\log p}{\lambda e^{\beta z_1}} \right]^{1/\gamma} \left[ \frac{\lambda e^{\beta z_0}}{-\log p} \right]^{1/\gamma} = \exp \left\{ \beta (z_0 - z_1) / \gamma \right\}. \]

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 \( \exp(-\beta_{\text{trt}}/\gamma) \).

∗Kantonsspital St. Gallen, Bedastrasse 1, 9007 St. Gallen (sarah.haile@gmail.com)
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,

coeff 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 coeff, 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 ?, 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

coeff

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>lambda</td>
<td>0.01853664</td>
<td>0.01898690</td>
</tr>
<tr>
<td>gamma</td>
<td>1.13014371</td>
<td>0.13844846</td>
</tr>
<tr>
<td>factor(stage)2</td>
<td>0.16692694</td>
<td>0.46112943</td>
</tr>
<tr>
<td>factor(stage)3</td>
<td>0.66289534</td>
<td>0.35550887</td>
</tr>
<tr>
<td>factor(stage)4</td>
<td>1.74502788</td>
<td>0.41476410</td>
</tr>
<tr>
<td>age</td>
<td>0.01973646</td>
<td>0.01424135</td>
</tr>
</tbody>
</table>

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

```r
> 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 (?survreg, Equations 12.2.13-18, with some modifications since R gives log σ).
2 WeibullDiag

The `WeibullDiag` function produces a diagnostic plot for Weibull AFT regression, similar to what is present (? 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.

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

![Weibull Regression Diagnostic Plot](image)