

## Chapter 3

# Parametric Survival Regression

**Definition 3.1.** In a *1D regression model*, the response variable  $Y$  is conditionally independent of the  $p \times 1$  vector of predictors  $\mathbf{x}$  given the sufficient predictor  $SP = h(\mathbf{x})$ , written

$$Y \perp\!\!\!\perp \mathbf{x} | SP \quad \text{or} \quad Y \perp\!\!\!\perp \mathbf{x} | h(\mathbf{x}), \quad (3.1)$$

where the real valued function  $h : \mathbb{R}^p \rightarrow \mathbb{R}$ . The estimated sufficient predictor  $ESP = \hat{h}(\mathbf{x})$ . An important special case is a model with a linear predictor  $h(\mathbf{x}) = \mathbf{x}^T \boldsymbol{\beta}$  where  $ESP = \mathbf{x}^T \hat{\boldsymbol{\beta}}$ .

An important class of parametric 1D regression models has  $Y | \mathbf{x} \sim D(\mathbf{x}^T \boldsymbol{\beta}, \boldsymbol{\gamma})$  where  $D$  is a parametric distribution that depends on the  $p \times 1$  vector of predictors  $\mathbf{x}$  only through  $SP = \mathbf{x}^T \boldsymbol{\beta}$ , and  $\boldsymbol{\gamma}$  is a  $q \times 1$  vector of parameters. Several important survival regression models, including Weibull regression and accelerated failure time models, have this form, and will be covered in this chapter. Weibull regression and Exponential regression are parametric proportional hazards regression models.

### 3.1 Weibull and Exponential Regression

**Definition 3.2.** For **parametric proportional hazards** regression models, the baseline function is parametric and the parameters are estimated via maximum likelihood. Then as a 1D regression model,  $SP = \boldsymbol{\beta}_P^T \mathbf{x}$ , and

$$h_{Y|SP}(t) \equiv h_{\mathbf{x}}(t) = \exp(\boldsymbol{\beta}_P^T \mathbf{x}) h_{0,P}(t) = \exp(SP) h_{0,P}(t)$$

where the parametric baseline function  $h_{0,P}$  depends on  $k$  unknown parameters but does not depend on the predictors  $\mathbf{x}$ . The survival function is

$$S_{\mathbf{x}}(t) \equiv S_{Y|SP}(t) = [S_{0,P}(t)]^{\exp(\boldsymbol{\beta}_P^T \mathbf{x})} = [S_{0,P}(t)]^{\exp(SP)}, \quad (3.2)$$

and

$$\hat{S}_{\mathbf{x}}(t) = [\hat{S}_{0,P}(t)]^{\exp(\hat{\boldsymbol{\beta}}_P^T \mathbf{x})} = [\hat{S}_{0,P}(t)]^{\exp(ESP)}. \quad (3.3)$$

The following univariate results will be useful for Exponential and Weibull regression. If  $Y$  has a Weibull distribution,  $Y \sim W(\gamma, \lambda)$ , then  $S_Y(t) = \exp(-\lambda t^\gamma)$  where  $t, \lambda$  and  $\gamma$  are positive. If  $\gamma = 1$ , then  $Y$  has an Exponential distribution,  $Y \sim EXP(\lambda)$  where  $E(Y) = 1/\lambda$ . See Examples 1.1 and 1.2. Now  $V$  has a smallest extreme value distribution,  $V \sim SEV(\theta, \sigma)$ , if

$$S_V(t) = P(V > t) = \exp\left(-\exp\left(\frac{t - \theta}{\sigma}\right)\right)$$

where  $\sigma > 0$  while  $t$  and  $\theta$  are real. If  $Z \sim SEV(0, 1)$ , then  $V = \theta + \sigma Z \sim SEV(\theta, \sigma)$  since the SEV distribution is a location scale family. Also,  $V = \log(Y) \sim SEV(\theta - \sigma \log(\lambda), \sigma = 1/\gamma)$ , and  $Y = e^V \sim W(\gamma = 1/\sigma, \lambda = e^{-\theta/\sigma})$ .

If  $Y_i$  follows a Weibull regression model, then  $\log(Y_i)$  follows an accelerated failure time (AFT) model:  $\log(Y_i) = \alpha + \boldsymbol{\beta}_A^T \mathbf{x}_i + \sigma e_i$  where the  $e_i$  are iid  $SEV(0, 1)$ , and  $\log(Y)|\mathbf{x} \sim SEV(\alpha + \boldsymbol{\beta}_A^T \mathbf{x}, \sigma)$ . See Section 3.2.

**Definition 3.3.** The **Weibull proportional hazards regression (WPH) model** or **Weibull regression model** is a parametric proportional hazards model with  $Y|\mathbf{x} \sim W(\gamma = 1/\sigma, \lambda_{\mathbf{x}})$  where

$$\lambda_{\mathbf{x}} = \exp\left[-\left(\frac{\alpha}{\sigma} + \frac{\boldsymbol{\beta}_A^T \mathbf{x}}{\sigma}\right)\right] = \lambda_0 \exp(\boldsymbol{\beta}_P^T \mathbf{x})$$

with  $\lambda_0 = \exp(-\alpha/\sigma)$  and  $\boldsymbol{\beta}_P = -\boldsymbol{\beta}_A/\sigma$ . Thus for  $t > 0$ ,  $P(Y > t|\mathbf{x}) =$

$$\begin{aligned} S_{\mathbf{x}}(t) &= \exp(-\lambda_{\mathbf{x}} t^\gamma) = \exp(-\lambda_0 \exp(\boldsymbol{\beta}_P^T \mathbf{x}) t^\gamma) = [\exp(-\lambda_0 t^\gamma)]^{\exp(\boldsymbol{\beta}_P^T \mathbf{x})} = \\ &= [S_{0,P}(t)]^{\exp(\boldsymbol{\beta}_P^T \mathbf{x})}. \end{aligned}$$

As a 1D regression model,  $Y|SP \sim W(\gamma, \lambda_0 \exp(SP))$ . Also,

$$h_i(t) = h_{Y_i|\mathbf{x}_i}(t) = h_{Y_i|\boldsymbol{\beta}_P^T \mathbf{x}_i}(t) = \exp(\boldsymbol{\beta}_P^T \mathbf{x}_i) h_0(t)$$

where  $h_0(t) = h_0(t|\boldsymbol{\theta}) = \lambda_0 \gamma t^{\gamma-1}$  is the Weibull **baseline function**. **Exponential regression** is the special case of Weibull regression where  $\sigma = 1$ . Hence  $Y|\mathbf{x} \sim W(1, \lambda_{\mathbf{x}}) \sim EXP(\lambda_{\mathbf{x}})$ .

Since Weibull regression and Exponential regression are proportional hazards regression models, the plots from Chapter 2 can be used to check the models. The Weibull proportional hazard model is valid iff the Weibull ac-

celerated failure time (AFT) model is valid. Similarly, the Exponential PH model is valid iff the Exponential AFT model is valid. Hence the following two plots are useful.

**Definition 3.4.** Let  $T_i = \min(Y_i, Z_i)$  be the censored survival times, and let  $\log(T_i) = \hat{\alpha} + \hat{\beta}_A^T \mathbf{x}_i + r_i$ . For accelerated failure time models, a **log censored response (LCR) plot** is a plot of  $\hat{\alpha} + \hat{\beta}_A^T \mathbf{x}_i$  versus  $\log(T_i)$  with plotting symbol 0 for censored cases and + for uncensored cases. The identity line with unit slope and zero intercept is added to the plot, and the vertical deviations from the identity line =  $r_i$ . Collett (2003b, p. 231) defines a standardized residual  $r_{Si} = r_i/\hat{\sigma}$ .

The least squares line based on the +’s could be added to the plot and should have slope not too far from 1, especially if  $\gamma \geq 1$  for the Weibull AFT. The plotted points should be linear with roughly constant variance. The censoring and long left tails of the smallest extreme value distribution make judging linearity and detecting outliers from the left tail difficult. Try to ignore the bottom of the plot where there are few cases when assessing linearity.

**Definition 3.5.** For parametric proportional hazards models, an **EE plot** is a plot of the parametric ESP  $\hat{\beta}_P^T \mathbf{x}$  versus the Cox semiparametric ESP  $\hat{\beta}_C^T \mathbf{x}$ .

If the parametric proportional hazards model is good, then the plotted points in the EE plot should track the identity line with unit slope and zero intercept. As  $n \rightarrow \infty$ , the correlation of the plotted points goes to 1 in probability for any finite interval, e.g., from the 1st percentile to the 99th percentile of  $\hat{\beta}_C^T \mathbf{x}$ . Lack of fit is suggested if the plotted points do not cluster tightly about the identity line.

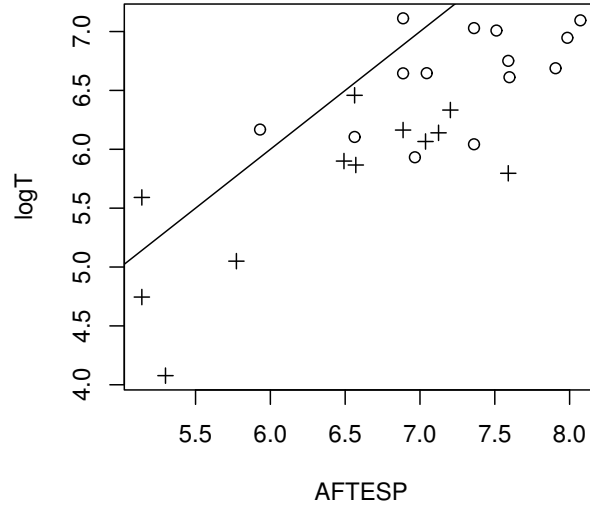
Software typically fits Exponential and Weibull regression models as accelerated failure time models:  $\log(Y_i) = \alpha + \beta_A^T \mathbf{x}_i + \sigma e_i$ . For the Exponential regression model,  $\sigma = 1$  and  $\beta_C = -\beta_A$ , and the Exponential EE plot is a plot of

$$ESPE = -\hat{\beta}_A^T \mathbf{x} \text{ versus } ESPC = \hat{\beta}_C^T \mathbf{x}.$$

For the Weibull regression model,  $\beta_C = -\beta_A/\sigma$ , and the Weibull EE plot is a plot of

$$ESPW = \frac{-1}{\hat{\sigma}} \hat{\beta}_A^T \mathbf{x} \text{ versus } ESPC = \hat{\beta}_C^T \mathbf{x}.$$

Suppose the plotted points cluster tightly about the identity line in the EE plot with  $\text{corr}(\hat{\beta}_C^T \mathbf{x}_i, \hat{\beta}_P^T \mathbf{x}_i) > 0.99$ . Thus  $\hat{\beta}_C^T \mathbf{x} \approx \hat{\beta}_P^T \mathbf{x}$  for the observed  $\mathbf{x}_i$ , and slicing on the Cox ESP is nearly the same as slicing on the parametric ESP. Make the slice survival plot for the Cox model and add the estimated parametric survival function (3.3) as crosses. If the parametric proportional



**Fig. 3.1** LCR Plot for Ovarian Cancer Data

hazards model holds, then (2.2) = (3.2). Thus if (2.3)  $\approx$  (3.3) for any  $\mathbf{x}_i$ , then  $S_{0,P}(t) \approx S_0(t)$ , (2.3)  $\approx$  (3.3) for all  $\mathbf{x}_i$ , and the parametric proportional hazards model is reasonable.

**Remark 3.1.** Checking parametric proportional hazards models has 3 steps: i) check that the proportional hazards assumption is reasonable, e.g. with the slice survival plot for the Cox model, ii) check that the parametric and semiparametric ESPs are approximately the same,  $\hat{\beta}_P^T \mathbf{x} \approx \hat{\beta}_C^T \mathbf{x}$  with the EE plot, and iii) using the slice survival plot, check that (2.3)  $\approx$  (3.3) for the  $\mathbf{x}$  used in each of the  $J$  slices. Since the Weibull proportional hazards model (Def. 3.3) is valid for  $(Y, \mathbf{x})$  if and only if the Weibull accelerated failure time model (Def. 3.7) is valid for  $(\log(Y), \mathbf{x})$ , the above procedure can be used to simultaneously check the goodness of fit of both models.

This technique avoids the mistake of comparing quantities from the semi-parametric and parametric proportional hazards models without checking that the proportional hazards assumption is reasonable. The slice survival plot for the Cox model is used because of the ease of making pointwise CI bands.

**Example 3.1.** The ovarian cancer data is from Collett (2003b, p. 187-190) and Edmunson et al. (1979). The response variable is the survival time

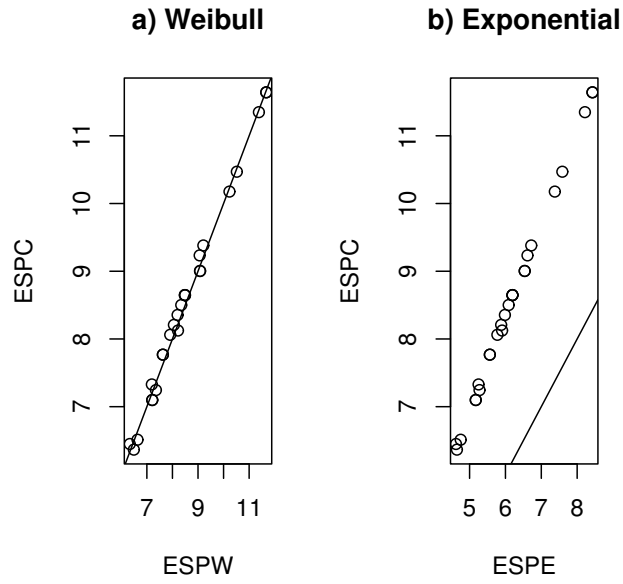


Fig. 3.2 EE Plots for Ovarian Cancer Data

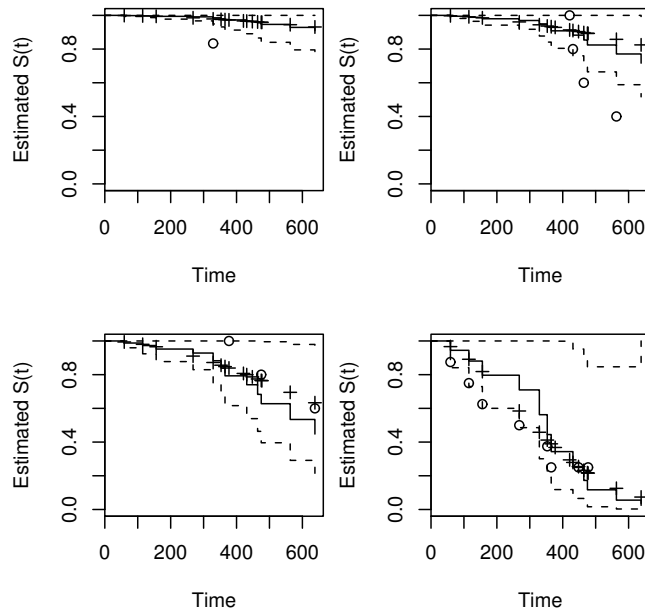


Fig. 3.3 Slice Survival Plots for Ovarian Cancer Data

of  $n = 26$  patients in days with predictors *age* in years and *treat* (1 for cyclophosphamide alone and 2 for cyclophosphamide combined with adriamycin). Figure 3.1 shows that most of the plotted points in the LCR plot for the ovarian cancer data are below the identity line. If a Weibull regression model is a good approximation to the data, then the plotted points in a narrow vertical slice centered at  $\hat{\alpha} + \hat{\beta}^T \mathbf{x} = w$  are approximately a censored sample from an  $SEV(w, \hat{\sigma})$  distribution. Figure 3.2 shows the Weibull and Exponential regression EE plots. Notice that the estimated risk scores from the Cox regression and Weibull regression are nearly the same with correlation = 0.997. The points from the Exponential regression do not cluster about the identity line. Hence Exponential regression should not be used. Figure 3.3 gives the slice survival plot for the Cox model with the Weibull survival function  $\hat{S}_{\mathbf{x}}(t) = \exp[-\exp(-\hat{\gamma}\hat{\beta}_A^T \mathbf{x}) \exp(-\hat{\gamma}\hat{\alpha}) t^{\hat{\gamma}}]$  represented by crosses where  $\hat{\gamma} = 1/\hat{\sigma}$ . Notice that the Weibull and Cox estimated survival functions are close and thus similar. Again the circles corresponding to the Kaplan Meier estimator are “close” to the Cox survival curves in that the circles do not fall very far outside the pointwise CI bands.

Output for the Weibull and Exponential regression models is shown below. The output is often from software for accelerated failure time models. The tests are the same for the parametric PH model and the equivalent AFT model, but for Weibull regression the ESP and confidence intervals tend to be for  $\hat{\beta}_A = (\beta_i)$ , which differs from  $\hat{\beta}_P$  by a constant. Output for AFT models will include an intercept  $\hat{\alpha}$  and an estimate of scale  $\hat{\sigma}$ . *SAS* and *R* give output for the AFT.

For *SAS* or *R*.

variable	Est.	SE	Est/SE	or $(Est/SE)^2$	pvalue for
intercept					
$x_1$	$\hat{\beta}_1$	$se(\hat{\beta}_1)$	$z_{o,1} = \hat{\beta}_1/se(\hat{\beta}_1)$	$X_{o,1}^2 = z_{o,1}^2$	Ho: $\beta_1 = 0$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_p$	$\hat{\beta}_p$	$se(\hat{\beta}_p)$	$z_{o,p} = \hat{\beta}_p/se(\hat{\beta}_p)$	$X_{o,p}^2 = z_{o,p}^2$	Ho: $\beta_p = 0$
scale or	log scale				
Weibull shape	or scale				

Output for the null model for SAS is shown below.  
log likelihood log L(none)

variable	Est. SE	Est/SE or $(Est/SE)^2$	pvalue
intercept			
scale			
Weibull shape			

For the full model, SAS will have Log Likelihood = log L(full).  
For the full model, R will have log L(full), log L(none) and  
chisq =  $[-2 \log L(\text{none})] - [-2 \log L(\text{full})]$  on p degrees of freedom with pvalue.

Replace full by reduced for the reduced model.

**The SAS and R log likelihood, log L, differ by a constant.**

```
SAS Log Likelihood = -29.7672 null model
variable      df Estimate  SE      chi square  pr > chisqu
intercept     1    7.1110   0.2927   590.12      < 0.0001
Weibull Scale 1   1225.4   358.7
Weibull Shape 1    1.1081   0.2810
```

```
SAS Log Likelihood = -29.1775 reduced model
variable      df Estimate  SE      chi square  pr > chisqu
intercept     1    7.3838   0.4370   285.45      < 0.0001
treat         1   -0.5593   0.5292    1.12       0.2906
Scale         1    0.8857   0.2227
Weibull Shape 1    1.1291   0.2840
```

```
SAS Log Likelihood = -20.5631 full model
variable      df Estimate  SE      chi square  pr > chisqu
intercept     1   11.5483   1.1970   93.07       < 0.0001
age           1   -0.0790   0.0198   15.97       < 0.0001
treat         1   -0.5615   0.3399    2.73       0.0986
Scale         1    0.5489   0.1291
Weibull Shape 1    1.8218   0.4286
```

```
R reduced model Value Std. Error      z      p
(Intercept)    7.384    0.437 16.895 4.87e-64
treat         -0.559    0.529 -1.057 2.91e-01
Log(scale)    -0.121    0.251 -0.483 6.29e-01
Scale= 0.886
Loglik(model)= -97.4  Loglik(intercept only)= -98
Chisq= 1.18 on 1 degrees of freedom, p= 0.28
```

```

R full model      Value Std. Error      z      p
(Intercept)      11.548      1.1970    9.65 5.04e-22
treat             -0.561      0.3399   -1.65 9.86e-02
age              -0.079      0.0198   -4.00 6.43e-05
Log(scale)       -0.600      0.2353   -2.55 1.08e-02
Scale= 0.549
Loglik(model)= -88.7   Loglik(intercept only)= -98
Chisq= 18.41 on 2 degrees of freedom, p= 1e-04

```

Shown above is output in symbols from and *SAS* and *R*. The estimated coefficient is  $\hat{\beta}_j$ . The Wald chi square =  $X_{0,j}^2$  while  $p$  and “pr > chisqu” are both p-values.

Inference from output is much like that for the Cox PH regression model. Find the ESP,  $h_0(t)$ , 95% CI for  $\beta_i$ , do a Wald test for  $H_0 : \beta_i = 0$ , do a likelihood ratio test (LRT) for  $H_0 : \boldsymbol{\beta} = \mathbf{0}$  versus  $H_A : \boldsymbol{\beta} \neq \mathbf{0}$ , and do a change in LRT for  $H_0$ : the reduced model is good versus  $H_A$ : use the full model. The Cox PH regression model used a PLRT and a change in PLRT.

Given  $\hat{\boldsymbol{\beta}}$  from output and given  $\boldsymbol{x}$ , be able to find ESP =  $\hat{\boldsymbol{\beta}}^T \boldsymbol{x} = \sum_{i=1}^p \hat{\beta}_i x_i = \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_p x_p$ .

A large sample 95% CI for  $\beta_j$  is  $\hat{\beta}_j \pm 1.96 \text{ se}(\hat{\beta}_j)$ .

#### 4 step Wald test of hypotheses:

- i) State the hypotheses  $H_0: \beta_j = 0$   $H_A: \beta_j \neq 0$ .
- ii) Find the test statistic  $z_{0,j} = \hat{\beta}_j / \text{se}(\hat{\beta}_j)$  or  $X_{0,j}^2 = z_{0,j}^2$  or obtain it from output.
- iii) The p-value =  $2P(Z < -|z_{0,j}|) = P(\chi_1^2 > X_{0,j}^2)$ . Find the p-value from output or use the standard normal table.
- iv) If  $p\text{val} < \delta$ , reject  $H_0$  and conclude that  $X_j$  is needed in the Weibull survival model given that the other  $p - 1$  predictors are in the model. If  $p\text{val} \geq \delta$ , fail to reject  $H_0$  and conclude that  $X_j$  is not needed in the Weibull survival model given that the other  $p - 1$  predictors are in the model.

The 4 step likelihood ratio test **LRT** is

i)  $H_0 : \boldsymbol{\beta} = \mathbf{0}$   $H_A : \boldsymbol{\beta} \neq \mathbf{0}$

ii) test statistic  $X^2(N|F) = [-2 \log L(\text{none})] - [-2 \log L(\text{full})]$  is often obtained from output.

iii) The p-value =  $P(\chi_p^2 > X^2(N|F))$  where  $\chi_p^2$  has a chi-square distribution with  $p$  degrees of freedom. The p-value is often obtained from output.

iv) Reject  $H_0$  if the p-value  $< \delta$  and conclude that there is a WPH survival relationship between  $Y$  and the predictors  $\boldsymbol{x}$ . If p-value  $\geq \delta$ , then fail to reject  $H_0$  and conclude that there is not a WPH survival relationship between  $Y$  and the predictors  $\boldsymbol{x}$ .

For the above test,  $X^2(N|F)$  is Chisq from *R*. Both *R* and *SAS* give  $\log L$ , but for *R*,  $\log L_R = \log L + d_R$  and for *SAS*,  $\log L_{SAS} = \log L + d_{SAS}$ . So



$\log L$  differs by a constant for  $R$  and  $SAS$ , but the constant cancels with subtraction.

Also note that there could be a PH survival relationship but not a WPH survival relationship. Check WPH assumptions before doing inference.

The 4 step **change in LR test** is

i)  $H_0$ : the reduced model is good  $H_A$ : use the full model

ii) test statistic  $X^2(R|F) = X^2(N|F) - X^2(N|R) = [-2 \log L(red)] - [-2 \log L(full)]$ .

iii) The p-value =  $P(\chi_{p-r}^2 > X^2(R|F))$  where  $\chi_{p-r}^2$  has a chi-square distribution with  $p - r$  degrees of freedom.

iv) Reject  $H_0$  if the p-value  $< \delta$  and conclude that the full model should be used. If p-value  $\geq \delta$ , then fail to reject  $H_0$  and conclude that the reduced model is good.

**Example 3.2.** Between points 1) and 2) in the summary Section 3.4, is output for the ovarian cancer data of Example 3.1. This output is also shown in this section.

a) Find ESP if treat = 1 and age = 60.

Solution:  $ESP = \hat{\beta}^T \mathbf{x} = -0.561(1) - 0.079(60) = -5.301$ .

b) Find a 95% CI for  $\beta_1$  corresponding to treat.

Solution: Using output for the  $R$  full model, the 95% CI is  $\hat{\beta}_1 \pm 1.96 se(\hat{\beta}_1) = -0.561 \pm 1.96(0.3399) = -0.561 \pm 0.662 = [-1.2272, 0.1052]$ .  $SAS$  and  $R$  output differs slightly.

c) Test  $\beta_1 = 0$  corresponding to treat.

Solution: i)  $H_0 : \beta_1 = 0$   $H_1 : \beta_1 \neq 0$

ii)  $Z_{01} = \frac{-0.561}{0.3399} = -1.6504$  or use output

or  $X_{01}^2 = Z_{01}^2 = 2.7241$  (2.73 from output)

iii) pval =  $2P(Z < 1.65) = 2(0.0495) = 0.099$  (0.0986 from output)

or pval =  $P(\chi_1^2 > 2.72)$

```
df | .100 .05
-----
1 | 2.71 3.84
```

so  $0.05 < \text{pval} < 0.100$

iv) Fail to reject  $H_0$ . Treatment is not needed in the Weibull survival model given age is in the model.

d) Test  $\beta_2 = 0$  corresponding to age.

Solution: i)  $H_0 : \beta_2 = 0$   $H_1 : \beta_2 \neq 0$

ii)  $Z_{02} = -4.00$

or  $X_{02}^2 = 15.97$

iii) pval = 0.0000643 or pval  $< 0.001$

iv) Reject  $H_0$ . Age is needed in the Weibull survival model given treat is in the model.

e) Test  $\beta = \mathbf{0}$ .

Solution: i)  $H_0 : \beta = \mathbf{0}$   $H_1 : \beta \neq \mathbf{0}$

ii)  $R$ :  $X^2(N|F) = 18.41$  or

$X^2(N|F) = [-2 \log(L(\text{none}))] - [-2 \log L(\text{Full})] = [-2(-98)] - [-2(-88.7)] = 196 - 177.4 = 18.6$  due to rounding

or  $SAS$ :  $X^2(N|F) = [-2(-29.7672)] - [-2(-20.5631)] = 59.5344 - 41.1262 = 18.4082$

iii)  $pval = P(\chi_2^2 > 18.41)$

```
df | .001
-----
2 | 13.82
```

so  $pval < 0.001$  (0.0001 from output)

iv) Reject  $H_0$ : there is a WPH survival relationship between time  $Y$  and the predictors age and treat.

f) Test whether the reduced model with treat is good.

Solution: i)  $H_0$ : the reduced model is good  $H_1$ : use the full model

ii)  $R$ :  $X^2(R|F) = X^2(N|F) - X^2(N|R) = 18.41 - 1.18 = 17.23$  or

$X^2(R|F) = [-2 \log(L(\text{Red}))] - [-2 \log L(\text{Full})] = [-2(-97.4)] - [-2(-88.7)] = 194.8 - 177.4 = 17.4$  due to rounding

$SAS$ :  $X^2(R|F) = [-2(-29.1775)] - [-2(-20.5631)] = 58.355 - 41.1262 = 17.2288$

iii)  $pval = P(\chi_1^2 > 17.23)$

```
df | .001
-----
1 | 10.83
```

so  $pval < 0.001$

iv) Reject  $H_0$ : use the full model.

**Warning:** Remark 2.1 also applies for the model in this chapter.

## 3.2 Accelerated Failure Time Models

**Definition 3.6.** For a parametric *accelerated failure time* model,

$$\log(Y_i) = \alpha + \beta_A^T \mathbf{x}_i + \sigma e_i \quad (3.4)$$

where the  $e_i$  are iid from a location scale family. Let  $SP = \beta_A^T \mathbf{x}$ . Then as a 1D regression model,  $\log(Y)|SP = \alpha + SP + e$ . The parameters are again estimated by maximum likelihood and the survival function is

$$S_{\mathbf{x}}(t) \equiv S_{Y|\mathbf{x}}(t) = S_0 \left( \frac{t}{\exp(\beta_A^T \mathbf{x})} \right),$$

and

$$\hat{S}_{\mathbf{x}}(t) = \hat{S}_0 \left( \frac{t}{\exp(\hat{\boldsymbol{\beta}}_A^T \mathbf{x})} \right)$$

where  $\hat{S}_0(t)$  depends on  $\hat{\alpha}$  and  $\hat{\sigma}$ .

For the AFT model,  $h_i(t) = h_{\mathbf{x}}(t) = e^{-SP} h_0(t/e^{SP})$  and  $S_i(t) = S_{\mathbf{x}}(t) = S_0(t/\exp(SP))$  where  $SP = \boldsymbol{\beta}_A^T \mathbf{x}$ . If  $S_{\mathbf{x}}(t_{\mathbf{x}}(\rho)) = 1 - \rho$  for  $0 < \rho < 1$ , then  $t_{\mathbf{x}}(\rho)$  is the  $\rho$ th percentile. For the accelerated failure time model,

$$t_{\mathbf{x}}(\rho) = t_0(\rho) \exp(\boldsymbol{\beta}_A^T \mathbf{x})$$

where  $t_0(\rho) = \exp(\sigma e_i(\rho) + \alpha)$  and  $S_{e_i}(e_i(\rho)) = P(e_i > e_i(\rho)) = 1 - \rho$ . Note that the estimated percentile ratio is free of  $\rho$ ,  $\hat{\sigma}$  and  $\hat{\alpha}$

$$\frac{\hat{t}_{\mathbf{x}_1}(\rho)}{\hat{t}_{\mathbf{x}_2}(\rho)} = \exp(\hat{\boldsymbol{\beta}}_A^T (\mathbf{x}_1 - \mathbf{x}_2)).$$

The *acceleration factor*  $= e^{-SP}$  and  $t_{0,\rho} = e^{-SP} t_{\mathbf{x},\rho}$ . The median survival times are related by  $t_{0,0.5} = e^{-SP} t_{\mathbf{x},0.5}$ . If  $e^{-SP} < 1$ , then the median survival time of  $\mathbf{x} >$  the median survival time of  $\mathbf{0}$ , a result that is good if the event is death, but bad if the event is time until recovery. Note that  $H_{\mathbf{x}}(t) = -\log S_{\mathbf{x}}(t) = -\log(S_0(t/e^{SP})) = H_0(t/e^{SP})$ .

**Remark 3.2.** Assume  $x_i > 0$ . Then  $\beta_i > 0$  increases  $\log(Y_i)$  and  $Y_i$ , while  $\beta_i < 0$  decreases  $\log(Y_i)$  and  $Y_i$ . For the Cox PH regression model,  $h_{\mathbf{x}}(t) = \exp(\boldsymbol{\beta}^T \mathbf{x}) h_0(t)$ . Hence  $\beta_i > 0$  increases hazard and decreases  $Y_i$ , while  $\beta_i < 0$  decreases hazard and increases  $Y_i$ . In the WPH model,  $\boldsymbol{\beta}_P = -\boldsymbol{\beta}_A/\sigma$ .

The LCR plot of Definition 3.4 is still useful for finding influential cases for AFT models. If the Weibull PH regression model holds for  $Y_i$ , then  $\log(Y_i) = \alpha + \boldsymbol{\beta}_A^T \mathbf{x}_i + \sigma e_i$  where  $e_i \sim SEV(0, 1)$ . Thus  $\log(Y)|\mathbf{x} \sim SEV(\alpha + \boldsymbol{\beta}_A^T \mathbf{x}, \sigma)$ , and the  $\log(Y_i)$  follows a parametric accelerated failure time model. Two other important AFTs are i) the lognormal AFT where  $\log(Y)|\mathbf{x} \sim N(\alpha + \boldsymbol{\beta}_A^T \mathbf{x}, \sigma^2)$  where the  $Y_i$  are lognormal and the  $e_i \sim N(0, 1)$  are normal, and ii) the loglogistic AFT where  $\log(Y)|\mathbf{x} \sim L(\alpha + \boldsymbol{\beta}_A^T \mathbf{x}, \sigma)$  where the  $Y_i$  are loglogistic and the  $e_i \sim L(0, 1)$  are logistic. For the loglogistic AFT,  $Y$  follows a proportional odds model.  $Y$  does not follow a proportional hazards regression model for the loglogistic and lognormal AFTs. The residuals  $r_i$  are the vertical deviations from the identity line in the LCR plot, and should behave like a censored sample from the distribution of  $\sigma e_i$ . Hence the  $r_i$  are like a censored sample from i) a  $SEV(0, \sigma)$  distribution for a Weibull AFT, ii) a  $N(0, \sigma^2)$  distribution for a lognormal AFT, and iii) a  $L(0, \sigma)$  distribution for a loglogistic distribution. The normal and logistic distributions are symmetric.

**Definition 3.7.** The *Weibull AFT* satisfies  $\log(Y)|(\alpha + \beta_A^T \mathbf{x}) \sim SEV(\alpha + \beta_A^T \mathbf{x}, \sigma)$ . Thus points in a narrow vertical slice about  $\hat{\alpha} + \hat{\beta}_A^T \mathbf{x} = w$  in the LCR plot are approximately a censored sample from an  $SEV(w, \hat{\sigma})$  distribution if the fitted model is a good approximation to the data. The *Exponential AFT* is the special case with  $\sigma = 1$ .

**Theorem 3.1.** Weibull regression models, including Exponential regression models, are the only models where  $Y$  follows a proportional hazards regression model and  $\log(Y)$  follows an accelerated failure time model.

Censoring causes the bulk of the data to be below the identity line in the LCR plot. For example, Hosmer and Lemeshow (1999, p. 226) state that for the Exponential regression model,  $\hat{\alpha}$  forces

$$\sum_{i=1}^n \delta_i = \sum_{i=1}^n \frac{T_i}{\exp(\hat{\alpha} + \hat{\beta}_A^T \mathbf{x}_i)}.$$

Hence  $\hat{T}_i = \exp(\hat{\alpha} + \hat{\beta}_A^T \mathbf{x}_i) \approx (n / \sum_{i=1}^n \delta_i) T_i$  (roughly). With no censoring, the bulk of the data will still be lower than the identity line if the  $e_i$  are left skewed as for the Weibull regression model where the  $e_i \sim SEV(0, 1)$ .

**Remark 3.3.** Since the Weibull proportional hazards model is valid for  $(Y, \mathbf{x})$  if and only if the Weibull accelerated failure time model is valid for  $(\log(Y), \mathbf{x})$ , fit the data using Cox regression. Then the graphical procedure described in Remark 3.1 can be used to simultaneously check the goodness of fit of both the Weibull PH and AFT models. Similarly, the Exponential proportional hazards model is valid for  $(Y, \mathbf{x})$  if and only if the Exponential accelerated failure time model is valid for  $(\log(Y), \mathbf{x})$ .

For Weibull and Exponential regression, instead of fitting a PH model, *R* and *SAS* fit an accelerated failure time model  $\log(Y_i) = \alpha + \beta_A^T \mathbf{x}_i + \sigma e_i$  where the  $e_i$  are iid from a smallest extreme value distribution. The Exponential AFT is the special case of the Weibull AFT with  $\sigma = 1$ . As in Definition 3.10,  $\lambda_0 = \exp(-\alpha/\sigma)$  and  $\beta_P = -\beta_A/\sigma$  where  $\beta_P$  is the vector of coefficients for the WPH model and  $\beta_A$  is the vector of coefficients for the Weibull AFT model. Since the AFT is parametric,  $\hat{\alpha}$  and  $\hat{\beta}_A$  are MLEs found from the censored data  $(T_i, \delta_i, \mathbf{x}_i)$ , not from  $(Y_i, \mathbf{x}_i)$ .

If the  $Y_i|\mathbf{x}_i$  are Weibull, the  $e_i$  are from a smallest extreme value distribution. The statement that “*the Weibull regression model is both a proportional hazards model and an accelerated failure time model*” means that the  $Y_i|\mathbf{x}_i$  follow a Weibull PH model while the  $\log(Y_i)|\mathbf{x}_i$  follow a Weibull AFT, although the  $\log(Y_i)$  are actually from a smallest extreme value distribution. If a Weibull or Exponential AFT is a useful model for the  $\log(Y_i)|\mathbf{x}_i$ , then the Weibull or Exponential PH model is a good approximation for the  $Y_i|\mathbf{x}_i$ . Hence to check the goodness of fit for the Weibull AFT, transform the Weibull

AFT into the Weibull PH model. Then use the LCR, EE and slice survival plots as in Example 3.10.

Inference for the AFT model is performed almost in the same way as for the WPH or Weibull AFT. See Section 3.1. But the conclusions change slightly if the AFT is not the Weibull AFT. Change (if necessary) “Weibull survival model” to the appropriate model, e.g. “lognormal survival model”. In the LRT, replace “WPH” by “AFT.” Given  $\hat{\beta} \equiv \hat{\beta}_A$  from output and given  $\mathbf{x}$ , know how to find  $\text{ESP} = \hat{\beta}^T \mathbf{x} = \sum_{i=1}^p \hat{\beta}_i x_i = \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_p x_p$ .

A large sample 95% CI for  $\beta_j$  is  $\hat{\beta}_j \pm 1.96 \text{ se}(\hat{\beta}_j)$ .

Know how to do the **4 step Wald test of hypotheses**:

- i) State the hypotheses  $H_0 : \beta_j = 0$   $H_1 : \beta_j \neq 0$ .
- ii) Find the test statistic  $z_{0,j} = \hat{\beta}_j / \text{se}(\hat{\beta}_j)$  or  $X_{0,j}^2 = z_{0,j}^2$  or obtain it from output.
- iii) The p-value =  $2P(Z < -|z_{0,j}|) = P(\chi_1^2 > X_{0,j}^2)$ . Find the p-value from output or use the standard normal table.
- iv) If p-value  $< \delta$ , reject  $H_0$  and conclude that  $X_j$  is needed in the Weibull survival model given that the other  $p - 1$  predictors are in the model. If p-value  $\geq \delta$ , fail to reject  $H_0$  and conclude that  $X_j$  is not needed in the Weibull survival model given that the other  $p - 1$  predictors are in the model.

Know how to do the 4 step likelihood ratio test **LRT**:

- i)  $H_0 : \beta = \mathbf{0}$   $H_A : \beta \neq \mathbf{0}$
- ii) test statistic  $X^2(N|F) = [-2 \log L(\text{none})] - [-2 \log L(\text{full})]$  is often obtained from output.
- iii) The p-value =  $P(\chi_p^2 > X^2(N|F))$  where  $\chi_p^2$  has a chi-square distribution with  $p$  degrees of freedom. The p-value is often obtained from output.
- iv) Reject  $H_0$  if the p-value  $< \delta$  and conclude that there is an AFT survival relationship between  $Y$  and the predictors  $\mathbf{x}$ . If p-value  $\geq \delta$ , then fail to reject  $H_0$  and conclude that there is not an AFT survival relationship between  $Y$  and the predictors  $\mathbf{x}$ .

Know how to do the 4 step **change in LR test**:

- i)  $H_0$ : the reduced model is good  $H_A$ : use the full model
- ii) test statistic  $X^2(R|F) = X^2(N|F) - X^2(N|R) = [-2 \log L(\text{red})] - [-2 \log L(\text{full})]$ .
- iii) The p-value =  $P(\chi_{p-r}^2 > X^2(R|F))$  where  $\chi_{p-r}^2$  has a chi-square distribution with  $p - r$  degrees of freedom.
- iv) Reject  $H_0$  if the p-value  $< \delta$  and conclude that the full model should be used. If p-value  $\geq \delta$ , then fail to reject  $H_0$  and conclude that the reduced model is good.

### 3.3 Variable Selection

Since the Weibull proportional hazards model is valid for  $(Y, \mathbf{x})$  if and only if the Weibull accelerated failure time model is valid for  $(\log(Y), \mathbf{x})$ , fit the data using Cox PH regression and perform variable selection such as forward selection, backward elimination, and relaxed lasso. Then fit each candidate submodel with WPH software and check the WPH assumptions. Transform the PH model to a Weibull AFT if the AFT is desired. The following chapter shows how to do inference after variable selection.

### 3.4 Summary

1) The **Weibull proportional hazards regression (WPH) model** is

$$h_i(t) = h_{Y_i|\mathbf{x}_i}(t) = h_{Y_i|\boldsymbol{\beta}_P^T \mathbf{x}_i}(t) = \exp(\boldsymbol{\beta}_P^T \mathbf{x}_i) h_0(t)$$

where  $h_0(t) = h_0(t|\boldsymbol{\theta}) = \lambda_0 \gamma t^{\gamma-1}$  is the **baseline function**. So  $Y|SP \sim W(\gamma, \lambda_0 \exp(SP))$ .

**Assume that the WPH model is appropriate.**

For SAS only.  
log likelihood log L(none)

variable	Est.	SE	Est/SE or $(Est/SE)^2$	pvalue
intercept				
scale				
Weibull shape				

For SAS or R

variable	Est.	SE	Est/SE	or $(Est/SE)^2$	pvalue for
intercept					
$x_1$	$\hat{\beta}_1$	$se(\hat{\beta}_1)$	$z_{o,1} = \hat{\beta}_1 / se(\hat{\beta}_1)$	$X_{o,1}^2 = z_{o,1}^2$	Ho: $\beta_1 = 0$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_p$	$\hat{\beta}_p$	$se(\hat{\beta}_p)$	$z_{o,p} = \hat{\beta}_p / se(\hat{\beta}_p)$	$X_{o,p}^2 = z_{o,p}^2$	Ho: $\beta_p = 0$
scale or Weibull shape	log scale	or scale			

For the full model, SAS will have Log Likelihood = log L(full).

For the full model, R will have log L(full), log L (none) and  $chisq = [-2 \log L(\text{none})] - [-2 \log L(\text{full})]$  on p degrees of freedom with pvalue

Replace full by reduced for the reduced model.

**The SAS and R log likelihood, log L, differ by a constant.**

```
SAS Log Likelihood = -29.7672 null model
variable      df Estimate SE      chi square pr > chisqu
intercept     1   7.1110  0.2927  590.12      < 0.0001
Weibull Scale 1  1225.4  358.7
Weibull Shape 1  1.1081  0.2810
```

```
SAS Log Likelihood = -29.1775 reduced model
variable      df Estimate SE      chi square pr > chisqu
intercept     1   7.3838  0.4370  285.45      < 0.0001
treat         1  -0.5593  0.5292   1.12      0.2906
Scale         1   0.8857  0.2227
Weibull Shape 1  1.1291  0.2840
```

```
SAS Log Likelihood = -20.5631 full model
variable      df Estimate SE      chi square pr > chisqu
intercept     1  11.5483  1.1970  93.07      < 0.0001
age           1  -0.0790  0.0198  15.97      < 0.0001
treat         1  -0.5615  0.3399   2.73      0.0986
Scale         1   0.5489  0.1291
Weibull Shape 1  1.8218  0.4286
```

```
R reduced model Value Std. Error      z      p
(Intercept)      7.384      0.437 16.895 4.87e-64
treat            -0.559      0.529 -1.057 2.91e-01
Log(scale)       -0.121      0.251 -0.483 6.29e-01
Scale= 0.886
Loglik(model)= -97.4 Loglik(intercept only)= -98
Chisq= 1.18 on 1 degrees of freedom, p= 0.28
```

```
R full model Value Std. Error      z      p
(Intercept)     11.548      1.1970  9.65 5.04e-22
treat            -0.561      0.3399 -1.65 9.86e-02
age              -0.079      0.0198 -4.00 6.43e-05
Log(scale)       -0.600      0.2353 -2.55 1.08e-02
Scale= 0.549
Loglik(model)= -88.7 Loglik(intercept only)= -98
Chisq= 18.41 on 2 degrees of freedom, p= 1e-04
```

Shown above is output in symbols from and *SAS* and *R*. The estimated coefficient is  $\hat{\beta}_j$ . The Wald chi square =  $X_{o,j}^2$  while  $p$  and “pr > chisqu” are both p-values.

2) Instead of fitting the WHP model of 1), R and SAS fit an accelerated failure time model  $\log(Y_i) = \alpha + \beta_A^T \mathbf{x}_i + \sigma \epsilon_i$  where  $\text{Var}(\epsilon_i) = 1$  and the  $\epsilon_i$  are iid from a smallest extreme value distribution. Also  $\beta_A \neq \beta_P$  from 1).

$\hat{\alpha}$  and  $\hat{\beta}$  are MLEs found from the censored data  $(T_i, \delta_i, \mathbf{x}_i)$  not from  $(Y_i, \mathbf{x}_i)$ .

3) Let  $\log(T_i) = \hat{\alpha} + \hat{\beta}_A^T \mathbf{x}_i + r_i$ . A *log censored response (LCR) plot* is a plot of  $\hat{\alpha} + \hat{\beta}_A^T \mathbf{x}_i$  vs  $\log(T_i)$  with plotting symbol 0 for censored cases and + for uncensored cases. The vertical deviations from the identity line =  $r_i$ . The least squares line based on the +’s can be added to the plot, and should have slope not too far from 1 for the Weibull AFT if  $\gamma \geq 1$ . The plotted points should be linear with roughly constant variance. The censoring and long left tails of the smallest extreme value distribution make judging linearity and detecting outliers from the left tail difficult. Try to ignore the bottom of the plot where there are few cases when assessing linearity.

4) Given  $\hat{\beta}$  from output and given  $\mathbf{x}$ , be able to find  $\text{ESP} = \hat{\beta}^T \mathbf{x} = \sum_{i=1}^p \hat{\beta}_i x_i = \hat{\beta}_1 x_1 + \dots + \hat{\beta}_p x_p$ .

5) A large sample 95% CI for  $\beta_j$  is  $\hat{\beta}_j \pm 1.96 \text{ se}(\hat{\beta}_j)$ .

#### 6) 4 step Wald test of hypotheses:

i) State the hypotheses  $H_o: \beta_j = 0$   $H_a: \beta_j \neq 0$ .

ii) Find the test statistic  $z_{0,j} = \hat{\beta}_j / \text{se}(\hat{\beta}_j)$  or  $X_{0,j}^2 = z_{0,j}^2$  or obtain it from output.

iii) The p-value =  $2P(Z < -|z_{0,j}|) = P(\chi_1^2 > X_{0,j}^2)$ . Find the p-value from output or use the standard normal table.

iv) If  $\text{pval} < \delta$ , reject  $H_o$  and conclude that  $X_j$  is needed in the Weibull survival model given that the other  $p - 1$  predictors are in the model. If  $\text{pval} \geq \delta$ , fail to reject  $H_o$  and conclude that  $X_j$  is not needed in the Weibull survival model given that the other  $p - 1$  predictors are in the model.

7) The 4 step likelihood ratio test **LRT** is

i)  $H_o: \beta = \mathbf{0}$   $H_A: \beta \neq \mathbf{0}$

ii) test statistic  $X^2(N|F) = [-2 \log L(\text{none})] - [-2 \log L(\text{full})]$  is often obtained from output.

iii) The p-value =  $P(\chi_p^2 > X^2(N|F))$  where  $\chi_p^2$  has a chi-square distribution with  $p$  degrees of freedom. The p-value is often obtained from output.

iv) Reject  $H_o$  if the p-value  $< \delta$  and conclude that there is a WPH survival relationship between  $Y$  and the predictors  $\mathbf{x}$ . If p-value  $\geq \delta$ , then fail to reject  $H_o$  and conclude that there is not a WPH survival relationship between  $Y$  and the predictors  $\mathbf{x}$ .

8) The 4 step **change in LR test** is

i)  $H_o$ : the reduced model is good  $H_A$ : use the full model

ii) test statistic  $X^2(R|F) = X^2(N|F) - X^2(N|R) = [-2 \log L(\text{red})] - [-2 \log L(\text{full})]$ .



iii) The p-value =  $P(\chi_{p-r}^2 > X^2(R|F))$  where  $\chi_{p-r}^2$  has a chi-square distribution with  $p - r$  degrees of freedom.

iv) Reject  $H_o$  if the p-value  $< \delta$  and conclude that the full model should be used. If p-value  $\geq \delta$ , then fail to reject  $H_o$  and conclude that the reduced model is good.

9) R and SAS programs do not have a variable selection option, but the WPH model is a PH model, so use SAS Cox PH variable selection to suggest good submodels. Then fit each candidate with WPH software and check the WPH assumptions.

10) The **accelerated failure time (AFT) model** has  $\log(Y_i) = \alpha + \beta_A' \mathbf{x}_i + \sigma e_i$  where the  $e_i$  are iid from a location scale family.

If the  $Y_i$  are Weibull, the  $e_i$  are from a smallest extreme value distribution. The Weibull regression model is often said to be “both a proportional hazards model and an accelerated failure time model.” Actually the  $Y_i$  follow a PH models and the  $\log(Y_i)$  follow an AFT model.

If the  $Y_i$  are lognormal, the  $e_i$  are normal.

If the  $Y_i$  are loglogistic, the  $e_i$  are logistic.

11) Still use the *log censored response (LCR) plot* of 42). The LCR plot is easier to use when the  $\epsilon_i$  are normal or logistic since these are symmetric distributions.

12) For the AFT model,  $h_i(t) = e^{-SP} h_o(t/e^{SP})$  and  $S_i(t) = S_0(t/\exp(SP))$ .

13) Inference for the AFT model is performed exactly in the same way as for the WPH or Weibull AFT. See points 43) – 47). But the conclusion change slightly if the AFT is not the Weibull AFT. In point 45, change (if necessary) “Weibull survival model” to the appropriate model, eg “lognormal survival model”. In point 46, change (if necessary) “WPH” to the appropriate model, eg “lognormal AFT”.

In principle, the slice survival plot can be made for parametric AFT models, but the programming may be difficult.

The loglogistic and lognormal AFT models are not PH models. The loglogistic AFT is a proportional odds model.

14) Let  $\beta_C$  correspond to the Cox regression and  $\beta_A$  correspond to the AFT. An EE plot is a plot of the parametric ESP vs a semiparametric ESP with the identity line added as a visual aid. The plotted points should follow the identity line with a correlation tending to 1.0 as  $n \rightarrow \infty$ .

15) For the Exponential regression model,  $\sigma = 1$ , and  $\beta_C = -\beta_A$ . The Exponential EE plot is a plot of  $-ESPE = -\hat{\beta}'_A \mathbf{x}$  vs  $ESPC = \hat{\beta}'_C \mathbf{x}$ .

16) For the Weibull regression model,  $\sigma = 1$ , and  $\beta_C = -\beta_A/\sigma$ . The Weibull EE plot is a plot of

$$-ESPW/\hat{\sigma} = -\frac{1}{\hat{\sigma}}\hat{\beta}'_A \mathbf{x} \quad \text{vs} \quad ESPC = \hat{\beta}'_C \mathbf{x}.$$

### 3.5 Complements

The Weibull PH regression model is the most widely used parametric PH regression model, but the Cox semiparametric PH regression model is used much more often for survival analysis. When the Weibull PH regression model holds, the parametric inference is slightly better than the Cox PH regression model inference, but the Cox PH regression model gives good results for many data sets where the Weibull PH regression model does not hold.

A Weibull stratified PH regression model can be used where a Weibull PH regression model with the same  $\beta$  is used for each of the  $J$  strata. Then the  $\alpha_j$  and  $\sigma_j$  depend on the strata for  $j = 1, \dots, J$ .

For survival regression plots, see Olive (2011). Inference after variable selection and prediction intervals will be covered in Chapter 4.

A *proportional odds (PO) regression model* has

$$\frac{S_{\mathbf{x}}(t)}{1 - S_{\mathbf{x}}(t)} = e^{SP} \frac{S_{\mathbf{0}}(t)}{1 - S_{\mathbf{0}}(t)}$$

where  $SP = \beta_{PO}^T \mathbf{x}$ . The loglogistic regression model is the only model where  $\log(Y)$  follows an AFT and  $Y$  follows a proportional odds regression model. For the loglogistic model,  $\beta_{PO} = \beta_A/\sigma$ .

For a proportional odds regression model, note that

$$\frac{S_{\mathbf{x}}(t)}{1 - S_{\mathbf{x}}(t)} = \frac{P(Y > t|\mathbf{x})}{1 - P(Y > t|\mathbf{x})} =$$

odds of survival beyond time  $t$ . Then the log odds ratio is

$$\log \left[ \frac{\left( \frac{S_{\mathbf{x}}(t)}{1 - S_{\mathbf{x}}(t)} \right)}{\left( \frac{S_{\mathbf{0}}(t)}{1 - S_{\mathbf{0}}(t)} \right)} \right] = \beta_{PO}^T \mathbf{x}.$$

Wei (1992) and Zeng and Lin (2007) give nonparametric methods for AFTs. These methods could be used to check a parametric AFT much like the Cox PH regression model can be used to check a parametric PH regression model like the Weibull PH regression model. Similarly, Bennett (1983) and Yang and Prentice (1999) give nonparametric methods for the proportional odds (PO) regression model, and these method could be used to check the parametric loglogistic PO regression model.

### 3.6 Problems

Problems with an asterisk \* are especially important.

**3.1.** Leemis (1995, p. 190, 205-6) gives data on  $n = 21$  leukemia patients taking the drug 6-MP. Suppose that the remission times given below follow an exponential ( $\lambda$ ) distribution.

6, 6, 6, 6+, 7, 9+, 10, 10+, 11+, 13, 16, 17+,  
19+, 20+, 22, 23, 25+, 32+, 32+, 34+, 35+

- Find  $\hat{\lambda}$ .
- Find a 95% CI for  $\lambda$ .

**3.2.** Suppose that the lifetimes of a certain brand of lightbulb follow an exponential ( $\lambda$ ) distribution. 20 light bulbs are tested for 1000 hours. The failure times are below.

71, 88, 254, 339, 372, 403, 498, 499, 593, 774, 935,  
1000+, 1000+, 1000+, 1000+, 1000+, 1000+, 1000+, 1000+, 1000+

- Find  $\hat{\lambda}$ .
- Find a 95% CI for  $\lambda$ .

**3.3.** The following output is from a Weibull Regression for the Allison (1995, p. 270) recidivism data. The response variable *week* is time in weeks until arrest after release from prison (right censored if week = 52). The 7 variables are *Fin* (1 for those who received financial aid, 0 else), *Age* at time of release, *Race* (1 if black, 0 else), *Wexp* (1 if inmate had full time work experience prior to conviction, 0 else), *Mar* (1 if married at time of release, 0 else), *Paro* (1 if released on parole, 0 else), *Prio* (the number of prior convictions).

- For the reduced model, find a 95% CI for  $\beta_1$ .
- Test whether the reduced model is good.

```
Output for Problem 3.3  Null Model
Log Likelihood -336.08436 Standard 95% Confidence Chi-
Parameter  DF Estimate Error          Limits      Square Pr>ChiSq
Intercept  1   4.8177  0.1079   4.6062   5.0291 1994.47 <.0001
Scale      1   0.7325  0.0661   0.6138   0.8742
Weib Scale 1 123.6771 13.3417 100.1072 152.7964
Weib Shape 1   1.3651  0.1232   1.1438   1.6293
```

```
Full Model Log Likelihood -319.3765238
Standard 95% Confidence Chi-
Parameter  DF Estimate Error          Limits      Square Pr>ChiSq
Intercept  1   3.9901  0.4191   3.1687   4.8115  90.65 <.0001
fin        1   0.2722  0.1380   0.0018   0.5426   3.89  0.0485
```

age	1	0.0407	0.0160	0.0093	0.0721	6.47	0.0110
race	1	-0.2248	0.2202	-0.6563	0.2067	1.04	0.3072
wexp	1	0.1066	0.1515	-0.1905	0.4036	0.49	0.4820
mar	1	0.3113	0.2733	-0.2244	0.8469	1.30	0.2547
paro	1	0.0588	0.1396	-0.2149	0.3325	0.18	0.6735
prio	1	-0.0658	0.0209	-0.1069	-0.0248	9.88	0.0017
Scale	1	0.7124	0.0634	0.5983	0.8482		
Weib. Shape	1	1.4037	0.1250	1.1789	1.6713		

Reduced Model Log Likelihood -321.5012378

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr>ChiSq
Intercept	1	3.7738	0.3581	3.0720	4.4755	111.08	<.0001
fin	1	0.2495	0.1372	-0.0194	0.5184	3.31	0.0690
age	1	0.0478	0.0154	0.0176	0.0779	9.66	0.0019
prio	1	-0.0698	0.0201	-0.1092	-0.0304	12.08	0.0005
Scale	1	0.7141	0.0637	0.5995	0.8506		
Weib. Shape	1	1.4004	0.1250	1.1756	1.6681		

Output for Problem 3.4

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr>ChiSq
Intercept	1	3.7738	0.3581	3.0720	4.4755	111.08	<.0001
fin	1	0.2495	0.1372	-0.0194	0.5184	3.31	0.0690
age	1	0.0478	0.0154	0.0176	0.0779	9.66	0.0019
prio	1	-0.0698	0.0201	-0.1092	-0.0304	12.08	0.0005
Scale	1	0.7141	0.0637	0.5995	0.8506		
Weibull Shape	1	1.4004	0.1250	1.1756	1.6681		

**3.4.** Above is output from a Weibull Regression for the Allison (1995, p. 270) recidivism data described in problem 3.3. The full model has 3 predictors, *fin*, *age* and *prio*.

a) Suppose that the log likelihood for the null model is  $-336.08436$ . Test whether  $\beta = \mathbf{0}$ .

b) Test whether  $\beta_1 = 0$ .

c) Test whether  $\beta_2 = 0$ .

Output for 3.5

	Value	Std. Error	z	p
(Intercept)	5.32632	0.66298	8.03	9.44e-16
age	-0.00891	0.00711	-1.25	0.210
sex	0.37019	0.12796	2.89	0.00382
ph.karno	0.00926	0.00446	2.08	0.0379
Log(scale)	-0.28085	0.06171	-4.55	5.33e-06

Scale= 0.755

Weibull distribution

Loglik(model)= -1138.7    Loglik(intercept only)= -1147.5

Chisq= 17.59 on 3 degrees of freedom, p= 0.00053

n=227 (1 observation deleted due to missingness)

**3.5.** A Weibull regression model was fit to the *R* lung data resulting in the above output.

- Test whether  $\beta = \mathbf{0}$ .
- Test whether  $\beta_1 = 0$ .
- Test whether  $\beta_2 = 0$ .
- Sketch the Weibull EE plot if the Weibull model is good.

Output for 3.6, n = 26

	coef	exp(coef)	se(coef)	z	p	full model
age	0.121	1.13	0.0484	2.500	0.012	
resid.ds	0.792	2.21	0.8078	0.980	0.330	
ecog.ps	0.087	1.09	0.6592	0.132	0.890	
Likelihood ratio test= 13.7 on 3 df,						p=0.00333

	coef	exp(coef)	se(coef)	z	p	reduced model
age	0.137	1.15	0.0474	2.9	0.0038	
Likelihood ratio test= 12.7 on 1 df,						p=0.000368

**3.6.** The *R* ovarian data gives survival times for patients with ovarian cancer. Predictors are *age* in years, *resid.ds* (residual disease present 1=no,2=yes), and *ecog.ps* (ECOG performance status: 1 is better than 2). A stratified proportional hazards model is fit where the stratification variable *rx* is the treatment group.

- Test whether  $\beta_3 = 0$ .
- Test whether  $\beta = \mathbf{0}$  for the full model.
- Test whether the reduced model is good.

**3.7.** The *R* lung cancer data has the *time* until death or censoring. *ph.ecog* = Ecog performance score 0-4, *pat.karno* = patient's assessment of their karno score and *wt.loss* = weight loss in last 6 months. A stratified proportional hazards model is used and stratification is on *sex*.

- Find the ESP and  $\hat{h}_i(t)$  if  $\mathbf{x} = (1.0, 80.0, 7.0)$  and *sex* = *F*.
- Find a 95% CI for  $\beta_2$ .
- Do a 4 step test for  $H_0 : \beta_2 = 0$ .
- Do a 4 step test for  $H_0 : \beta_3 = 0$ .

e) *R* output says Likelihood ratio test=22.8.  
Do a 4 step test for  $H_0 : \beta = \mathbf{0}$ .

```
output for f)
      coef exp(coef) se(coef)      z      p
age      0.01444      1.01 0.010508  1.374 0.17
meal.cal -0.00016      1.00 0.000240 -0.666 0.51
```

```
Likelihood ratio test=2.97 on 2 df, p=0.227 n=181
(47 observations deleted due to missingness)
```

f) Now the SPH model uses the predictors *age* and *meal.cal* = calories consumed at meals excluding beverages and snacks.  
Do a 4 step test for  $H_0 : \beta = \mathbf{0}$ .

### R Problems

**3.8.** This problem considers the ovarian data from Collett (2003, p. 344-346).

a) Obtain the *R* code for 3.8 from (<http://parker.ad.siu.edu/Olive/survhw.txt>). Click on the left screen then hit *Enter*. Copy and paste both the output. (It should be very similar to that on Section 3.4 between points 1) and 2.) Also copy and paste the plot into *Word*.

b) The plot is a log censored response plot. The top line is the identity line and the bottom line the least squares line. Is the slope of the least squares line near 1?

**3.9.** 16.54: Use the source commands near the top of (<http://parker.ad.siu.edu/Olive/survhw.txt>) to get *survpack* into *R*. The programs *phdata*, *weyp* and *wregsim* will be used.

The program *wregsim* generates Weibull proportional hazards regression data with baseline hazard function  $h_0(t) = kt^{k-1}$ .

a) Type the command *wregsim(k=1)* 5 times (or use the “up arrow” after typing the command once). This gives 5 simulated Weibull regression data sets with  $k = 1$ . Hence the Weibull regression is also an exponential regression. Include the last plot in *Word*.

b) Type the command *wregsim(k=5)* 5 times. To judge linearity, ignore the cases on the bottom of the plot with low density (points with log(time) less than -2). (These tend to be censored cases because time  $Y = W^{1/k}$  where  $W \sim EXP(\lambda = \exp(SP))$  where  $E(W) = 1/\lambda$ .  $Z \sim EXP(.1)$  has mean 10 and if  $Z_i < Y_i$  then  $Z_i$  is usually very small.) Do the plots seem linear ignoring the cases on the bottom of the plot? Do not include the plot.

c) Type the command *wregsim(k=0.5)* 5 times. (Now censored cases tend to be large because time  $Y = W^{1/k} = W^2$  where  $W \sim EXP(\lambda)$ .  $Z \sim EXP(.1)$  has mean 10 and if  $Z_i < Y_i$  then  $Y_i > 10$ , usually.) Do the plots seem linear (ignoring cases on the bottom of the plot)? (The plot is linear if

it is roughly box shaped or ellipsoidal, possibly ignoring some of the points with  $\log(\text{time}) < -9$ . Since the error distribution is left skewed, most of the plotted points will fall below the identity line, even if the plot is linear.) Do not include the plot.

**3.10.** 16.55: This problem considers the ovarian data from Collett (2003, pp. 189, 344-346).

a) Obtain the *R* code for 3.10a from (<http://parker.ad.siu.edu/Olive/survhw.txt>). Copy and paste the plot into *Word*.

b) Now obtain the *R* code for 3.10b and put the plot into *Word*.

c) Can the Exponential regression model be used or should the more complicated Weibull regression model be used?

**3.11.** Copy and paste the two source commands from the top of (<http://parker.ad.siu.edu/Olive/survhw.txt>) to get programs *phdata* and *wregsim2* into *R*.

Make the left window small by moving the cursor to the lower right corner of the window, then hold the right mouse button down and drag the window to the left.

The program *wregsim2* generates Weibull proportional hazards regression data with baseline hazard function  $h_0(t) = kt^{k-1}$ .

a) Type the command *wregsim2*( $n=10, k=1$ ) 5 times (or use the “up arrow” after typing the command once). This gives 5 simulated Weibull regression data sets with  $k = 1$ . Increase  $n$  by 10 until the plotted points cluster tightly about the identity line in at least 4 out of 5 times. How big is  $n$ ?

b) Type the command *wregsim2*( $n=10, k=5$ ) 5 times. Increase  $n$  by 10 until the plotted points cluster tightly about the identity line in at least 4 out of 5 times. How big is  $n$ ?

c) Type the command *wregsim2*( $n=10, k=0.5$ ) 5 times. Increase  $n$  by 10 until the plotted points cluster tightly about the identity line in at least 4 out of 5 times. How big is  $n$ ?

**3.12.** If necessary copy and paste the two source commands as done for Problem 3.11 to get programs *phdata* and *wregsim3* into *R*.

Make the left window small by moving the cursor to the lower right corner of the window, then hold the right mouse button down and drag the window to the left.

The program *wregsim3* generates Weibull proportional hazards regression data with baseline hazard function  $h_0(t) = kt^{k-1}$ . This is also an AFT model with  $\alpha = 0$ ,  $\beta' = -(1/k, \dots, 1/k)$  and  $\sigma = 1/k$ . The program generate 100 Weibull AFT data sets and for each run  $i$  computes  $\hat{\alpha}_i$ ,  $\hat{\beta}_i$  and  $\hat{\sigma}_i$ . Then the averages are reported. Want  $\text{mnint} \approx 0$ ,  $\text{mnccoef} \approx -(1/k, \dots, 1/k)$  and  $\text{mnscale} \approx 1/k$ .

a) Make a table (by hand) with headers

n	k	mnint	mncoef	mnscale
---	---	-------	--------	---------

Fill in the table for  $n = 20, k = 1; n = 100, k = 1; n = 200, k = 1; n = 20, k = 5; n = 100, k = 5; n = 200, k = 5; n = 20, k = 0.5; n = 100, k = 0.5; n = 200, k = 0.5$  by using the commands `wregsim3(n=20, k=1)`, ..., `wregsim3(n=200, k=0.5)`.

b) Are the estimators close to parameters  $\alpha, \beta$  and  $\sigma$  for  $n = 20$ ? How about for  $n = 100$ ?

**3.13.** If necessary copy and paste the two source commands as done for problem 3.11 to get programs `wphsim` and `swhat` into *R*. Type the command `wphsim(n=999)` to make a slice survival plot based on the WPH survival function. Are the KM curve and Weibull estimated survival function close for the plot in the bottom right corner? Include the plot in *Word*.

**3.14.** The *R* lung cancer data has the *time* until death or censoring and *status* = 0 for censored and 1 for uncensored. Then the covariates are *age*, *sex* = 1 for M and 2 for F, *ph.ecog* = Ecog performance score 0-4, *ph.karno* = a competitor to *ph.ecog*, *pat.karno* = patient's assessment of their karno score, *meal.cal* = calories consumed at meals excluding beverages and snacks and *wt.loss* = weight loss in last 6 months. The *R* output will use a stratified proportional hazards model that is stratified on *sex* with variables *ph.ecog*, *pat.karno* and *wt.loss*.

a) Copy and paste commands from (<http://parker.ad.siu.edu/Olive/survhw.txt>) for this problem into *R*. Click on the left window and hit *Enter*. Include the plot in *Word*. Also include the *R* output in *Word*.

b) Test whether  $\beta = \mathbf{0}$ .

c) Based on the plot, do females or males appear to have better survival rates?

### SAS Problem

**3.15.** This problem considers the ovarian data from Collett (2003, p. 344-346).

a) Obtain the SAS program for 3.15 from (<http://parker.ad.siu.edu/Olive/survhw.txt>). Print the output. (It should be very similar to that on Section 3.4 between points 1) and 2).)

b) Find the ESP if  $age = 40$  and  $treat\ 1 = 1$ . (Comment: treatment takes on 2 levels so only one indicator is needed. SAS output includes a 2nd indicator *treat 2* but its coefficient is  $\hat{\beta}_3 = 0$  and hence can be ignored. In general if the category takes on J levels, SAS will give nonzero output for the first J - 1 levels and a line of 0s for the Jth level. This means level J was omitted and the line of 0s should be ignored.)

c) Give a 95% CI for  $\beta_1$  corresponding to age from output and the CI using the formula.



d) Give a 95% CI for  $\beta_2$  corresponding to treat 1 from output and the CI using the formula.

e) If the model statement in the SAS program is changed to  
model survtime\*status(0)=;

then the null model is fit and the SAS output says Log Likelihood  $-29.76723997$ .

Test  $\beta = \mathbf{0}$  with the LR test.

(Hint: The full model log likelihood  $\log(L) = -20.56313339$ . Want  $-2 \log(L)$  for both the full and null models for the LR test.)

f) Suppose the reduced model does not include *treat*. Then SAS output says Log Likelihood  $-21.7830$ . Test whether the reduced model is good.

(Hint: The log likelihood for the full model is  $\log(L) = -20.56313339$ . Want  $-2 \log(L)$  for the full and reduced models for the change in LR test.)