Survival Data Analysis (BIOS:7210)
Breheny

Assignment 7
Due: Thursday, October 31

1. As we saw in last week's homework, the score test for the hypothesis $H_0 : \beta_j = 0$ is based on

$$u_j \sqrt{(\mathbf{I}^{-1})_{jj}} \dot{\sim} \mathrm{N}(0,1),$$

where the score $\mathbf{u}$ and information $\mathbf{I}$ are both evaluated at $\beta_j = 0$ and where $\boldsymbol{\beta}_{-j}$ maximizes $\ell(\boldsymbol{\beta}_{-j}|\beta_j = 0)$.

   (a) Suppose we fit an exponential regression model with $\mathbf{X}$ as a design matrix, thereby obtaining $\widehat{\boldsymbol{\beta}}$, $\mathbf{W}$, and $\mathbb{V}(\widehat{\boldsymbol{\beta}}) = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}$. We are considering adding a new variable, β^*. In terms of $\mathbf{X}$, $\mathbf{W}$, and $\mathbb{V}(\widehat{\boldsymbol{\beta}})$ from the original fit, give expressions for $(\mathbf{I}^{-1})_{**}$ and the score statistic for testing $H_0 : \beta^* = 0$. These expressions should not contain any matrix inverses.

   (b) Fit an exponential regression model to the `pbc` data, with `stage` and `bili` as explanatory variables. Carry out score tests for the significance of `trt`, `hepato`, and `ascites`. Note that carrying out these tests can be performed without ever fitting a model with any of the new terms present.

2. Suppose that we fit the Weibull regression model stated at the bottom of slide 19 on the "Weibull distribution" notes to obtain estimates $\hat{\alpha}$, $\hat{\boldsymbol{\beta}}^*$, and $\hat{\sigma}$. Based on these estimates, we are interested in predicting the survival distribution of the time-to-event T for an individual with covariates $\mathbf{x}_i$. State what distribution T follows, along with the parameters of that distribution, in terms of $\hat{\alpha}$, $\hat{\boldsymbol{\beta}}^*$, and $\hat{\sigma}$.

3. The problem re-visits the question of a treatment-age interaction in the GVHD data that we also examined in Assignment 6. As in Assignment 6, censor the times on study at 60 days (for any subject still at risk on day 60). Now, however, fit a Weibull AFT model to the data, again including Age, Group, and the Group by Age interaction.

   (a) Is there significant evidence that the Weibull model provides a superior fit to the data than an exponential model?

   (b) Provide an estimate along with a 95% confidence interval for the Weibull shape parameter γ. Provide some explanation for how you arrived at this interval.

   (c) Consider the question of estimating the treatment effect for an individual of age a. Let $\widehat{\boldsymbol{\beta}}$ denote the vector of parameter estimates yielded by the AFT model you fit, with covariance matrix $\mathbb{V}\widehat{\boldsymbol{\beta}}$. There is no explicit parameter for the treatment effect at age a. However, the treatment effect can be constructed as a linear combination of the parameters in the model. I.e., there exists a contrast Δ_a such that $\Delta_a^T\boldsymbol{\beta}$ is equal to the treatment effect at age a. What is Δ_a? Also, provide an expression for the Wald 95% confidence interval for $\Delta_a^T\boldsymbol{\beta}$ in terms of Δ_a, $\widehat{\boldsymbol{\beta}}$, and $\mathbb{V}\widehat{\boldsymbol{\beta}}$.

   (d) Use your result from (c) to construct a plot of the age-specific treatment effect versus age. Include a gray shaded band to represent the confidence intervals. If you have not done this before, a confidence band can be plotted using the `polygon` function using the following code:

```
polygon(c(a, rev(a)), c(lwr, rev(upr)), col='gray85', border=NA)
```

where `lwr` and `upr` are vectors containing the lower and upper bounds of the interval. In one or two sentences, summarize the main implications of the plot as far as what it says about the treatment benefits of MTX+CSP compared to MTX alone.

(e) Plot the estimated survival functions for four representative patients: (1) a 10 year-old treated with MTX alone (2) a 10 year-old treated with MTX+CSP (3) a 30 year-old treated with MTX alone (4) a 30 year-old treated with MTX+CSP. Provide a single plot with the four survival curves overlaid, along with a legend or some other annotation to indicate which line belongs to which subject.

4. This problem consists of deriving the information matrix for the Weibull AFT model and using it to construct a confidence interval for σ. For parts (e) and (f), you will construct a confidence interval using data from a real AFT model and compare your results to what you get from the `survival` package. For these parts, use the Age by Treatment interaction model that we fit in the previous problem (i.e., for the GVHD data censored at 60 days, the model with Age, Group, and the Group by Age interaction).

(a) Using the book's notation (sort of; actually, the book defines $\mathbf{a}$ to be the negative of how we're defining it here), $\mathbf{a}$ is an $n \times 1$ vector with elements

$$a_i = \frac{\partial g}{\partial w_i},$$

and $\mathbf{A}$ is an $n \times n$ matrix with elements

$$A_{ij} = -\frac{\partial^2 g}{\partial w_i \partial w_j}.$$

Recall that g was defined in the notes as follows:

$$g(\mathbf{w}) = \sum_i \{d_i \log \lambda(w_i) + \log S(w_i)\},$$

where $\lambda(\cdot)$ and $S(\cdot)$ are the hazard and survival function for W. Derive $\mathbf{a}$ and $\mathbf{A}$ for the Weibull AFT model.

(b) Derive $\partial^2 \ell / \partial \boldsymbol{\beta}^2 |_{\hat{\boldsymbol{\theta}}}$; express your answer in terms of σ, $\mathbf{X}$, and $\mathbf{A}$. Note that for (b)-(d), some terms are equal to zero when evaluated at the MLE.

(c) Derive $\partial^2 \ell / \partial \boldsymbol{\beta} \partial \sigma |_{\hat{\boldsymbol{\theta}}}$; express your answer in terms of σ, $\mathbf{X}$, $\mathbf{A}$, and $\mathbf{w}$.

(d) Derive $\partial^2 \ell / \partial \sigma^2 |_{\hat{\boldsymbol{\theta}}}$; express your answer in terms of σ, $\mathbf{A}$, $\mathbf{a}$, and $\mathbf{w}$.

(e) Using your answers from (b)-(d), calculate a confidence interval for σ in the model described above. Note that you can obtain $\hat{\boldsymbol{\eta}}$ from `fit$linear.predictors` and $\mathbf{X}$ from `model.matrix(fit)`.

(f) Use your confidence interval from (e) to obtain a confidence interval for the Weibull shape parameter γ. How does it compare to the answer from the `survival` package (i.e., your answer to part b of the previous problem)?