

Survival Data Analysis (BIOS 7210)
Breheny

Assignment 9
Due Thursday, November 16

1. This question concerns Wald and likelihood ratio tests for the global null hypothesis $\beta = \mathbf{0}$ in the Cox model. Fit the following model to the GVHD data:

```
fit <- coxph(Surv(Time, Status) ~ Group + LAF + Age, gvhd)
```

- (a) In terms of the MLE $\hat{\beta}$ and the information, \mathbf{I} , of the fitted model, provide a mathematical expression for the global Wald test statistic, and say what distribution it follows under the null.
 - (b) Using `coef(fit)` and `vcov(fit)`, calculate the global Wald test statistic and verify that it equals the value in the output of `summary(fit)`.
 - (c) Under the assumption of no ties among times on study, provide a mathematical expression for the log-likelihood of the null Cox model. The expression should be in terms of $\{d_i\}$ and $\{r_i\}$, the time-on-study rank for subject i , only (i.e., it should not contain π or w terms; summations are allowed, however).
 - (d) Using `Status`, `rank(Time)`, and `logLik(fit)`, calculate the global likelihood ratio test statistic and verify that it equals the value in the output of `summary(fit)`. NOTE: Because there are ties in the GVHD data, your answer will not match the output from `summary(fit)` exactly. Feel free to artificially break ties by jittering survival time to verify that they are exactly equal, but you don't have to.
 - (e) Repeat part (d), only this time calculate the log likelihood of the null model under the Breslow approximation, and carry out a likelihood ratio test comparing it to `logLik(fit)`. Verify that it equals the output of `summary(fit)` exactly when `ties='breslow'` is specified.
2. In class, we derived the relationship

$$\lambda_{ij} = 1 - (1 - \lambda_{0j})^{w_i}.$$

Show that in the continuous case, this relationship still yields the usual Cox proportional hazards assumption in the limit. In other words, show that

$$\lim_{\epsilon \rightarrow 0} \frac{1 - [1 - \{\Lambda_0(t + \epsilon) - \Lambda_0(t)\}]^{w_i}}{\epsilon} = \lambda_0(t)w_i.$$

3. (a) Under independent censoring, the i th subject contributes a $\lambda_i(t_i)^{d_i} S_i(t_i)$ term towards the likelihood, where λ_i and S_i are the hazard and survival functions for the i th subject. Show that, under the proportional hazards assumption, the i th contribution towards the log-likelihood is given by

$$\ell_i(\eta_i) = d_i \eta_i - e^{\eta_i} \Lambda_0(t_i),$$

up to a constant not depending on η_i .

(b) Show that

$$\max_{\eta_i} \ell_i(\eta_i) = -d_i \log \Lambda_0(t_i) - d_i.$$

(c) In class, we used $\tilde{\ell}_i$ to denote the term from (b) and expressed the deviance residual in terms of $2(\tilde{\ell}_i - \ell_i)$. Show that

$$\max_{\eta_i} \ell_i(\eta_i) - \ell_i(\hat{\eta}_i) = -\hat{m}_i - d_i \log(d_i - \hat{m}_i);$$

in other words, the deviance residual can be written in terms of the martingale residual, \hat{m}_i , and the failure indicator d_i .

4. The course website contains a data set (Byar1980) from a randomized controlled trial in which men with stage 3 or stage 4 prostate cancer were assigned to either a therapy involving estrogen (at various doses) or a placebo. The outcome is all-cause mortality (time until death from any cause). The purpose of this problem is to assess the functional form of a few key covariates. For each covariate, (i) briefly describe why you chose the functional form that you did, (ii) present the estimate of its effect, along with confidence interval(s), and (iii) in one sentence, describe the nature of the covariate's effect on survival. For (ii), if the effect is linear, a hazard ratio will suffice. If the effect is nonlinear, a plot will be necessary.

- (a) Serum hemoglobin (**hg**)
- (b) Tumor size (**sz**)
- (c) Age (**Age**)
- (d) Serum prostatic acid phosphatase (**ap**)