is has t/am Questions.

Survival Analysis - Winter 2019

Assignment 3

The assignment is due Wednesday March 20th at **2:10pm**.

1. The data below are from a randomized trial studying the role of male circumcision in HIV prevention in Uganda. The events are incident HIV infections during the first 24 months of follow-up since the randomization and the intervention is male circumcision.

	Control arm ($Z = 0$)	Intervention arm ($Z = 1$)
Participants	$N_0 = 2430$	$N_1 = 2387$
Incident events	$D_0 = 45$	$D_1 = 22$
Person-years	$Y_0 = 3391.8$	$Y_1 = 3352.4$

A regression model fitted to the data gave the following output:

```
Call:
glm(formula = d ~ z + offset(log(y)), family = poisson(link = "log"))

Deviance Residuals:
[1]  0  0

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -4.3225     0.1491 -28.996  < 2e-16 ***
z            -0.7039     0.2601  -2.706  0.00681 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance:  7.7920e+00  on 1  degrees of freedom
Residual deviance: -7.9936e-15  on 0  degrees of freedom
AIC: 14.585

Number of Fisher Scoring iterations: 3
```

Specify the regression model corresponding to the above output in terms of a baseline hazard parameter and a regression coefficient. Write the likelihood expression for these parameters, and apply the profile likelihood method to eliminate the baseline hazard parameter. Use the obtained profile likelihood function to estimate the regression coefficient and compare the result to the above output.

Olli Saarela, Dalla Lana School of Public Health, 155 College Street, 6th floor, Toronto, Ontario M5T 3M7, Canada. E-mail: olli.saarela@utoronto.ca

2. In the Veterans administration lung cancer trial data set, males with advanced inoperable lung cancer were randomized to either a standard or experimental chemotherapy. In addition to treatment status, the data set includes a number of prognostic factors. The data set can be found from the R survival package.

   (a) Visual checking of the proportionality of hazard functions in the context of Cox models is generally complicated by the fact that hazard functions are difficult to estimate non-parametrically and that proportionality of two curves is difficult to assess visually. However, a graphical check can be based on the result $\log[-\log S_1(t)] - \log[-\log S_0(t)] = \beta$ which applies under the proportional hazard model for survival functions $S_1(t)$ and $S_0(t)$ corresponding to levels 1 and 0 of a binary covariate. This result implies that the log-log-transformed survival curves are parallel under the proportional hazards assumption; parallelity is easier to assess visually compared to proportionality, and the survival curves can be estimated using the Kaplan-Meier method. Prove the above result.

   (b) Present and interpret a log-log plot to check the proportionality of the treatment effect in the Veteran lung cancer trial.

   (c) If we fit a Cox model to estimate the treatment effect, the survival functions in the two treatment groups could be estimated by $\hat{S}_0(t) = \exp\{-\hat{\Lambda}_0(t)\}$ and $\hat{S}_1(t) = \exp\{-\hat{\Lambda}_0(t)\exp(\hat{\beta})\}$, where $\hat{\Lambda}_0(t)$ is the Breslow estimate for the cumulative baseline hazard. Present a log-log plot of such survival functions and interpret it. Why this plot is not useful for checking the proportionality assumption?

3. We might be interested in visually checking the proportionality assumption of the treatment effect in a Cox model that has been adjusted for prognostic factors in the dataset. For this purpose, we could fit a stratified Cox model, which is of the form

$$\lambda_i(t) = \lambda_{z_i}(t)\exp(\beta' x_i),$$

   where $z_i \in \{0, 1\}$ is the treatment arm indicator and x_i are the other covariates in the model. This allows for separate baseline hazards for the two treatment groups. In R, such a model would be fitted by not including the treatment arm indicator as a covariate, but instead entering it through a `strata` term. Separate Breslow estimates for the two baseline hazards can now be obtained to compare the baseline survival functions. Present and interpret a log-log plot of the baseline survival functions to check the proportionality assumption of the treatment effect in the adjusted model.

4. Fit a Cox model to estimate the treatment effect while adjusting for the prognostic factors. For this model,

   (a) Investigate the appropriateness of the linearity assumptions made on the effects of the continuous covariates using appropriate residual plots.

   (b) Check also the presence of potential influential observations.

   (c) Investigate the appropriateness of the proportinal hazards assumptions through appropriate residual plots, and statistical tests. Report the tests for both the residual-time correlations, and covariate-time interactions added to the Cox model (the latter can be implemented using the `tt` argument of the `coxph` function).

5. (a) Calculate the martingale residuals for the model in Q4 using the formulas given in the lecture material. Show your algorithm. Verify that the residuals match to those given by the `residuals` function.

   (b) Show (generally) that in the sample martingale residuals sum to zero.