Homework 4

Due Friday 12 April 2019

Read all questions carefully before answering. You may work in small groups of no more than 3 individuals and turn in a single assignment (and everyone in the group will receive the same grade). Work through the entire assignment individually first, then come together to discuss and collaborate. Please maintain numbering on sub-questions, type your responses, and **please keep answers brief.** Also, please do not submit computer code or unformatted output.

When assessing interaction via the likelihood ratio test use a significance criteria of 5%.

Data preliminaries

Load in required packages and the dataset (the name of the relevant data frame is `finaldata.stroke`):

```
require(survival)
require(multcomp)
require(lmtest)
require(car)

load("framingham_hw4.RData")
```

Dose-response modeling

Returning to our analysis of smoking and stroke in the Framingham cohort, we note that it hasn't been very satisfying to consider smoking as a binary variable. We will examine the dose-response association between smoking (measured in cigarettes per day) and incident stroke. Using the dataset provided, consider this association using simple linear coding, and spline coding with boundaries at: 20, 40, and 60 cigarettes per day.

Tasks:

- Referring to the example from the lecture notes, code variables for restricted quadratic splines for cigarettes per day (`cigpday`).
- Fit two Cox proportional hazards models with the follow-up variable `timestrk` and outcome indicator `stroke`, both adjusted for age (linear in log-hazard).
  - Code cigarettes per day as linear in the log-hazard. Store these results in `stroke.linear`.
  - Code cigarettes per day as a restricted quadratic spline. Store these results in `stroke.rqs`.
- Using the code below, plot predicted *risk score* ($\exp(\mathbf{x}\hat{\beta})$) at median age for each model, side-by-side (for your own review–you won't be turning this in):

```
# Create a new dataset where age is fixed to its median value:
# (will use this for calculating predicted values)
finaldata.stroke.new <- finaldata.stroke
```

```
finaldata.stroke.new$age <- median(finaldata.stroke$age)

par(mfrow=c(1,2)) # Plot matrix with 1 row and 2 columns:
plot(finaldata.stroke.new$cigpday, predict(stroke.linear, type="risk",
                                           newdata=finaldata.stroke.new),
     main=expression(paste("Linear cig/day")), log="y",
     xlab="Cig/day", ylab=expression(paste("exp(X*",hat(beta),")")))
plot(finaldata.stroke.new$cigpday, predict(stroke.rqs, type="risk",
                                           newdata=finaldata.stroke.new),
     main=expression(paste("RQS cig/day")), log="y",
     xlab="Cig/day", ylab=expression(paste("exp(X*",hat(beta),")")))
par(mfrow=c(1,1)) # Reset plot matrix to 1 row/column
```

- For the spline model, estimate hazard ratios and 95% confidence intervals to compare 1) 30 cigarettes per day vs. 10 cigarettes per day and 2) 50 cigarettes per day vs. 10 cigarettes per day. *Hint: follow the procedure outlined in the lecture notes for Dose-response modeling under "Reporting Results".*
- Compare the simple linear coding and restricted quadratic spline coding with a likelihood ratio test:

```
anova(stroke.linear, stroke.rqs)
```

Interaction and Effect Modification

Use the dataset provided to explore the relationship between BMI at baseline (`bmi_cat`) and time to stroke (`timestrk`, `stroke`) according to sex (`male` indicator variable) in the Framingham cohort. For the purposes of this assignment you may assume any necessary assumptions for proportional hazards models are met.

Tasks:

- Fit two Cox models (both adjusted for age (linear) and education (categorical)):
  - One with BMI-sex interaction (+ main effects); store the results in `stroke.bmi.ixn.sex`, and
  - One with main effects for BMI-sex (no interactions); store the results in `stroke.bmi.adj`.
- Assess multiplicative interaction between BMI and sex with the likelihood ratio test:

```
# Multiplicative
summary(stroke.bmi.ixn.sex) # Identify positions of the coefficients
anova(stroke.bmi.adj, stroke.bmi.ixn.sex) # LR p-value for multiplicative IXN
```

- Estimate the common-referent hazard ratios (using normal weight females as reference), <u>and</u> the straum-specific hazard ratios (e.g. the hazard ratios for overweight and obese BMI vs. normal weight among males). Use the `multcomp::glht` command. *Question 4 asks you to show part of your work for this task.*
- Calculate the RERI for additive interactionsb between 1) overweight BMI-sex,a nd 2) obese BMI-sex, and using the delta method (`car::deltaMethod`), corresponding 95% confidence intervals. Store

these results in `reri.sex.bmi25` and `reri.sex.bmi30`, respectively *Question 5 asks you to show part of your work for this task.*

Questions

Dose-response modeling

1. Write out **both** Cox models that we estimated for the simple linear (Model 0) and restricted quadratic spline (Model 1) coded version of cigarettes per day in terms of the variables above, and general parameters (e.g. β coefficients). *Hint: you may express the spline terms succintly and define them below the equation (see notes).* **(10 points)**
2. Consider the likelihood ratio test that compares that simple linear and restricted quadratic spline models. In words, describe what this is testing. **Specifically in terms of the model parameters** (i.e. β coefficients), what are the null and alternative hypotheses that are being tested (use formal notation)? What expression of cigarette smoking would you choose based on this test? **(10 points)**
3. Using the results from the spline model, report the hazard ratios and 95% confidence intervals comparing: **(10 points)**
   a. 30 cigarettes per day vs. 10 cigarettes per day.
   b. 50 cigarettes per day vs. 10 cigarettes per day.

Interaction and Effect Modification

4. Write out the expressions for the following in terms of the model parameters and show the corresponding coefficient vector (**k**, above) for the contrast of interest. **(Hint: for each of these, write out the two equations for relevant log-hazard functions to help show your work.)**. **(30 points)**
   a. The HR comparing obese (BMI $\geq$ 30) vs. normal weight (18.5 $\leq$ BMI $<$ 25.0) among females.
   b. The HR comparing obese (BMI $\geq$ 30) vs. normal weight (18.5 $\leq$ BMI $<$ 25.0) among males
   c. The HR comparing obese (BMI $\geq$ 30) males vs. normal weight (18.5 $\leq$ BMI $<$ 25.0) females.
5. Write out the expression for the RERI for the obesity (BMI $\geq$ 30)-sex interaction in terms of the hazard in each exposure group. Define/interpret each term in the equation and express each term using coefficients from the above regression model. **(10 points)**
6. Complete Table 1, being sure to include point estimates **and** confidence intervals, where appropriate, as well as the relevant parts of the footnotes. **(15 points)**
7. What are your conclusions regarding effect modification on the additive and multiplicative scales? Your answer should include individual interpretations of interaction on each scale. **(5 points)**
8. What assumptions regarding confounding have you made in assessing effect modification? If you were instead interested in the causal interaction between BMI and these variables would you have to consider any other relationships? If so, what? **(10 points)**

Table 1: Adjusted hazards ratio estimates and 95% CI of the association between baseline BMI status and mortality. The Framingham Cohort Study. 1948-1972, Framingham, MA.

BMI	Female	Male (single ref)	Male (stratum-specific)
18.5-24.9	1.		1.
25.0-29.9			
$\geq$ 30.0			

All hazard ratios adjusted for ______________________.

Effect modification on additive scale (95% CI): RERI male vs. female/25.0-29.9 vs. 18.5-24.9: _____; RERI male vs. female/ ≥30.0 vs. 18.5-24.9: _____.

Effect modification on multiplicative scale (95% CI): Male vs. female/25.0-29.9 vs. 18.5-24.9: _____ ; Male vs. female/≥30.0 vs. 18.5-24.9: _____; Likelihood ratio test p-value: _____.