Course EPIB-634: Survival Analysis & Related Topics [Winter 2007]   Assignment 6
[See SAS/Stata/R programs provided under Resources in www.epi.mcgill.ca/hanley/c681/cox

1. Sharper and Fairer Comparisons: effect of sexual activity on the longevity of male fruitflies
[Limit analysis to fruitflies with 1 partner .. the effect is obvious in those with 8]

When we first analyzed these data, PE, now on McGill faculty, argued that thorax size cannot be used as a predictor/explanatory variable since fruitflies who die young may not be fully grown, i.e., it is also an "intermediate" variable. Later, NK (now on faculty elsewhere) had studied entomology & assured us that fruitflies do not grow longer after birth; i.e., thorax length is not time- (age)-dependent!

Q3 (optional) in assignment 5 focused on the analysis of these data by the traditional ("identity link, Gaussian variation about the x-specific means") regression model -- the one fitted using proc *REG* in SAS, or *regress* in Stata, or *lm* in R, in the first half of course 621. This model, focuses on estimates of the (absolute) difference in *mean longevity* of sexually active flies (index category) relative to sexually inactive flies (reference category), ignoring, or taking acoount of, other covariates.

The longevity of the 2 groups can also be compared by survival analysis methods that instead yield hazard ratios (a ratio > 1 implies *shorter* survival). There is no requirement that some of the lifetimes have to be censored.

i   Use p h models to obtain "crude" & adjusted hazard ratios contrasting mortality rates of active vs. inactive flies (treat thorax size as continuous; ignore sleep for now)

ii   Write out the model you would have used if approaching these data from the vantage of Q3 of assignment 5.

Use logs to write the p h model so right hand side has the same additive form as in one for Q3 of ass'nt 5.

iii What role might the variable SLEEP have in this analysis? Based on the data, does it seem to be influential/relevant?

2   Clayton &Hills, Ch 30; Exercises 30.1 and 30.2 (attached)

3   Fit Cox's proportional hazards model directly, for a small dataset with 1 binary 'X' variate, namely the data on the longevity of the 10 fruitflies (5 sexually active and 5 inactive shown in Figure 1, and its legend .. same one as in assignment 4). We also illustrate the use of the stratified Cox model, where instead of 'modelling' the effect of thorax size, we 'match' on categories of thorax size.

(i) The objective is to find -- directly -- the (Maximum Likelihood) "B_hat" (the log of the "active vs. inactive" hazard ratio) in the Cox proportional hazards model that ignores thorax size. i.e.,

$$h[\,t \mid active\,] = h_0[t] \times \exp[\,B \times Active\,]$$

without resorting to specialized software such as coxph in R, phreg in SAS, or stcox in Stata.

If you wish to know ahead of time what the 'right answer' for the HR estimate from the Cox model is, run the coxph program on the 10 datapoints (remember the actual longevity values themselves don't matter; it is just their ordering that enters the (partial) Likelihood). (Incidentally, you will find that the M-H estimate of the rate ratio was 2.46, not far from the estimate from the Cox model).

```
longevity = c(30,35,40,45,50,65,70,75,80,90);
dead      = c( 1, 1, 1, 1, 1, 1, 1, 1, 1, 1);
active    = c( 1, 0, 1, 0, 1, 1, 0, 1, 0, 0);
```

To obtain B_hat directly, use the R code provided under Resources, and change the range of b so as to zero-in on B_hat , the value of b where the likelihood is maximized. Note that with large datasets, the likelihood -- a product -- is a very small number, and so the log.likelihood (a sum of logs) is usually tracked instead. If you don't have R, you could set up, and find the maximum of, the Likelihood in an Excel spreadsheet.

(ii) Refer to Fig 3, and legend, where the objective is to compare 'like' active with 'like' inactive fruitflies, i.e., smaller with smaller, larger with larger, but to arrive at a single hazard ratio estimate (you used these data in assignment 4 in a stratified version of the Log-Rank test). Adapt the n_0 and n_1 vectors, and plot the likelihood function, so as to obtain the B_hat (and by exponentiation, the HR) for the *stratified Cox model.*

(iii) Instead of matching on thorax, create a vector Large.thorax=c(...) containing ten 0/1 indicators of "Large.thorax". Add this variate to the model in (i) and obtain -- via the coxph module-- the (adjusted) estimate of the B coefficient (and the corresponding HR) for the Active vs. Inactive contrast.

Sexual activity and the lifespan of male fruitflies

ABSTRACT: A cost of increased reproduction in terms of reduced longevity has been shown for female fruitflies, but not for males. The flies used were an outbred stock. [In a randomized trial] sexual activity was manipulated by supplying individual males with one or eight receptive virgin females per day. The longevity of these males was compared with that of two control types. The first control consisted of two sets of individual males kept with one or eight newly inseminated females. Newly inseminated females will not usually remate for at least two days, and thus served as a control for any effect of competition with the male for food or space. The second control was a set of individual males kept with no females. There were 25 males in each of the five groups, which were treated identically in number of anaesthetizations (using CO2) and provision of fresh food medium.

SOURCE:
Figure 2 in the article "Sexual Activity and the Lifespan of Male Fruitflies" by Linda Partridge and Marion Farquhar. Nature, 294, 580-581, 1981.

```
Variable      Description
--------      -----------
ID            Serial No. (1-25) within each group of 25
              (the order in which data points were abstracted)

PARTNERS      Number of companions (0, 1 or 8)

ACTIVE          Type of companion
                0: newly pregnant female
                1: virgin female
                9: not applicable (when PARTNERS=0)

LONGEVITY     Lifespan, in days

THORAX        Length of thorax, in mm (0.xx)

SLEEP         Percentage of each day spent sleeping
```

NOTES: "Compliance" of the males in the two experimental groups was documented as follows: On two days per week throughout the life of each experimental male, the females that had been supplied as virgins to that male were kept and examined for fertile eggs. The insemination rate declined from approximately 7 females/day at age one week to just under 2/day at age eight weeks in the males supplied with eight virgin females per day, and from just under 1/day at age one week to approximately 0.6/day at age eight weeks in the males supplied with one virgin female per day. These `compliance' data were not supplied for individual males, but the authors say that "There were no significant differences between the individual males within each experimental group."

If interested, see also: Hanley, J. A. (1983), "Appropriate Uses of Multivariate Analysis," *Annual Review of Public Health*, 4, 155-180. (on c697 website under other Resources) and *Nature* article -- and especially the original Figure from the Nature article (on c622 website under Datasets)

Life cycle of Drosophila: The drosophila egg is about half a millimeter long. It takes about one day after fertilisation for the embryo to develop and hatch into a worm-like larva. The larva eats and grows continuously, moulting one day, two days, and four days after hatching (first, second and third instars). After two days as a third instar larva, it moults one more time to form an immobile pupa. *Over the next four days, the body is completely remodelled to give the adult winged form, which then hatches from the pupal case and is fertile within about 12 hours*. (timing is for 25°C; at 18°, development takes twice as long.)   [http://www.ceolas.org/fly/intro.html]