Course EPIB-634: Survival Analysis & Related Topics [Winter 2007]
Assignment 5

1 Refer to Chapter 22 of Clayton &Hills (chapter was attached to material handed out on March 2, and is also available in www.epi.mcgill.ca/hanley/c634/rates). Use Poisson regression to fit the (how-many-parameters?) model in Table 22.5 (or in symbols in Table 22.3) to the data in Table 22.6, thereby obtaining the regression coefficients in Table 22.8. From these fitted coefficients, work out the entries for the "baseline" or "corner" rate, and the 3 multipliers, shown in Table 22.7. From these, derive the table of 6 fitted rates and thus the 6 fitted counts. How close are the fitted and observed numbers of cases? Given that the dataset consists of 6 independent counts, how many degrees of freedom are left for assessing the fit of the model?

The data and fitting statements using R are available in the www.epi.mcgill.ca/hanley/c634/rates page, along with the corresponding statements for SAS and Stata.

2 The structure of the data in Table 3 of the report on the Uganda circumcision RCT is similar to the one you have just analyzed in Q1 above. JH fitted the 'second-smallest' model to these data (see page 5 of March 2 handout on regression models for 'event rate' data). *The data and fitting statements using R are available in the www.epi.mcgill.ca/hanley/c634/rates page,*

(i) Fit the *smallest* model, where you assume neither intervention nor follow-up interval has any impact on the incidence rate. Look at the fit (e.g. the fitted vs. observed numbers of cases, and the 'deviance') and comment.

(ii) Does the (somewhat larger) model shown in Tables 22.3/22.5 produce a significantly better fit to the Uganda data than the 2 parameter one reported in the handout? *[when dealing with Gaussian variation, it is common to use a partial F test as a way to compare larger/smaller models. In generalized linear models, one criterion commonly used is the likelihood ratio or deviance].*

(iii) What about a model where "incidence is same over all follow-up periods in control group and in first 6 months in intervention group, but falls and stays constant thereafter in the intervention group .. "?

"**Warning**" {small print, should be larger!} : question (iii) should not be taken as an endorsement of 'post hoc model selection' in the analysis of data from an RCT. One should state the analysis model up front in such applications. Indeed, in many applications in epidemiology and medicine, "building" models, and trying to find the "best" one, seem silly, unless we are in the prediction business -- and even then the 'quest for the best' is often overdone/unrealistic. This dataset is merely being used here to get you familiar with translating "expected patterns, expressed in words" into their "statistical/mathematical equivalents" that can be ultimately be expressed as regression equations !

(iv) Using the reported incidence in each period in each group, and the relationship between incidence and risk, calculated what the 2-year risk (cumulative incidence) would be if uncircumcised / circumcised. Compare these with the cumulative probabilities reported from the Kaplan-Meier analyses.

(v) Explain to the editors (and their referees who also missed this) why the reported 'cumulative incidence' values for the total 0-24 months follow-up period are not correct. (Table 3).

3. [OPTIONAL, helpful if wish to 'see' how multiple regression 'corrects for' confounding]

"Sharper and Fairer" Comparisons: effect of sexual activity on the longevity of male fruitflies (cf below) *[data and programs in www.epi.mcgill.ca/hanley/c681/cox]* [limit analysis to fruitflies with 1 partner; the effect is obvious in those with 8]

When we first analyzed this dataset, student PE, now on McGill faculty, argued that thorax size cannot be used as a predictor or explanatory variable since fruitflies who die young may not be fully grown, i.e., it is also an "intermediate" variable. Later, student 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!

a *Use a regression module such as PROC REG in SAS, or lm in R, or regress in Stata, to estimate the (absolute) difference in mean longevity of sexually active flies (index category) relative to sexually inactive flies (reference category), ignoring other covariates. Is this difference (i) substantial? (ii) statistically significant at the conventional alpha=0.05 level?*

b *How well did the randomization work? How different are the thorax lengths of the active and inactive flies? Is this difference "statistically" significant? Is statistical significance a non-issue here anyway? Explain.*

c *If -- other things being equal -- flies 0.01 mm larger live on average 1 day longer, how much of a longevity "advantage" would the active flies have as a result of their larger average thorax size? On this basis, how much lower is the mean longevity of active than inactive flies if "adjusted" for the difference in thorax size?*

d *Instead of using the "out of the air" value of 1 day/ 0.01 mm, use multiple regression to simultaneously estimate the additional mean days/mm and the decrease in days associated with (due to) activity i.e., fit the model:* $average[longevity \mid thorax\ type] = B_0 + B_{thorax} \times thorax + B_{active} \times active$

e *Verify that if you correct/adjust the comparison in (a) using the fitted* b_{thorax} *from (d) and the thorax difference in (b), you arrive at the* b_{active} *obtained in (d). Hint: cf JH notes on confounding (attached)*

f The p-value for the activity contrast in (d) is smaller (& the associated CI narrower) than the corresponding one in (a). One reason is that the larger adjusted estimate of the effect (the numerator of t-test on *adjusted* difference); another is the *smaller* SE of the estimated effect (denominator of t-test). *Why is the SE of the estimated longevity difference from analysis (d) smaller?*

In the next assignment, the longevity of the 2 groups will also be compared by survival analysis methods.

4. [OPTIONAL, example of *additive and multiplicative models for 'rates'* in an familiar but non-epi context]

See the attached Liste de prix / Price list for the Lefoft2 condominiums (2005 prices). *The data are also in an electronic file in www.epi.mcgill.ca/hanley/c634/rates*

The data-analysis objective is to figure out what formula the realtor might have used for the $/pi^2 'rate' as an (increasing) function of the floor. Other unit-specific factors, such as location on floor, probably entered into it too, but ignore them for now; likewise, ignore the first (ground) floor units. Here are two 'rate' possibilities:

(i) a rate that was linear for every extra floor above the second e.g.

rate = $/pi^2 = \$200 + \$5 for every extra floor ('additive')

(ii) a 'multiplicative' one, where rate was more progressive, e.g.

(a) rate = $/pi^2 = $\$200 \times (1 + 0.025)^{\text{extra floor}}$

(b) rate = $/pi^2 = $\$200 \times e^{0.025 \times \text{extra floor}}$

These are analogs of 'simple-' and 'compound-interest' models

(i) Using model (i), write out the formula for the (average) price of a unit as a function of floor and square footage, and use the dataset to estimate the "base", i.e. "floor-2", rate (what C&H call the "corner") and "per extra floor" amount in 'rate model' (i).

(ii) Using model (ii b), write out the formula for the (average) price of a unit as a function of floor and square footage, and use the dataset to estimate the corresponding coefficients in 'rate model' (ii b).

Compare the parameter estimates, and fits, of the 2 models. Work in $/pi^2 scale, not the log[$/pi^2] one!

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 (in www.epi.mcgill.ca/hanley/reprints or 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]