Solutions, Biostat III exam, November 14, 2008

1.

(a) (2 marks) From model 1 we can obtain an estimate of the effect of sex. For which other variables is this estimate adjusted?

Answer: The model is adjusted for calendar period, age at diagosis and time since diagnosis (the underlying time scale).

(b) (2 marks) Based on model 2, what is the estimated mortality rate ratio comparing papillary to follicular tumours for patients in the oldest age group (aged 70+ at diagnosis)? In other words, what is the effect of histology for the oldest age group? You do not need to comment on statistical significance.

Answer: The point estimate is 0.7097. The effect is assumed to be the same for all age groups (the fitted model is the main effects model).

(c) (2 marks) From model 1 we see that there is evidence that females experience lower DTC mortality than males. Is there evidence that this difference may be explained (fully or partly) by differences in the distribution of histological type between males and females (e.g., that females are more likely to be diagnosed with the less aggressive histological type)? Refer to the output from models 1 and/or 2 to support your answer.

Answer: No, the estimated hazard ratio changes from 0.58 to 0.59 when adjusted for histology. That is, there is no evidence that the effect of sex is confounded by histology.

(d) (4 marks) Is it possible, using results from models 1 and/or 2, to assess whether the effect of histology is modified by sex? If yes, comment on the magnitude of effect modification and whether it is statistically significant. If no, state how you would assess this.

Answer: No, this is not possible. In order to evaluate if there is effect modification we need to fit a model containing an interaction between histology and sex.

(e) (3 marks) Interpret the estimated hazard ratio for the variable labelled `_Iagegrp_60`, including a comment on statistical significance.

Answer: Patients with follicular cancer, aged 60-69 years at diagnosis, experience an 18.3 times higher mortality due to DTC compared to patients with follicular cancer aged 0-39 years at diagnosis.

The difference is statistically significant which can be seen either from:

- The z-statistic + p-value
- The confidence interval does not contain 1

(f) (3 marks) What do we mean (conceptually, not mathematically), when we state that an effect is 'statistically significant'? If a result is statistically significant does it mean there is a 'real' or 'true' association?

Answer: Statistical significance means that the observed differences are greater than what would be expected due to chance. That is, we do not believe the observed difference is due to chance. This does not necessarily imply that the association is 'true' since the difference may be due to bias or confounding.

(g) (3 marks) Based on model 3, what is the estimated mortality rate ratio (2 decimal places are sufficient) comparing papillary to follicular tumours for each age group? You do not need to present confidence intervals or comment on statistical significance.

Answer:

Agegroup	Mortality rate ratio
0–39	0.50
40–49	$0.50 \times 1.03 = 0.51$
50–59	$0.50 \times 1.04 = 0.52$
60–69	$0.50 \times 1.16 = 0.58$
70+	$0.50 \times 2.11 = 1.05$

(h) (3 marks) Is there evidence of a statistically significant interaction between histology and age group? If you choose to perform a formal hypothesis test you should state the null hypothesis, alternative hypothesis, value of the test statistic, assumed distribution of the test statistic under the null hypothesis, and a comment on statistical significance.

Answer: The formal approach to test whether there is need to include an interation term between histology and age group in the model is to conduct a likelihood ratio test.

The null hypothesis to be tested is:

$$H_0 : \text{All interaction effects are 0.}$$

against

$$H_A : \text{At least one of the interaction effects are non-zero.}$$

The test statistic is defined as

$$-2 \times (\text{log likelihood (model 2)} - \text{log likelihood (model 3)})$$

and has a χ^2 distribution under H_0. Note that the degrees of freedom is determined by how many 'extra' parateters we need to estimate in the model containing the interaction compared to the main effects model.

We thus calculate the difference in log likelihoods between model 2 (log likelihood: -7922.8) and model 3 (log likelihood: -7936.2) and get -13.4. This is mulitplied by -2 to get 26.8 which is our test statistic.

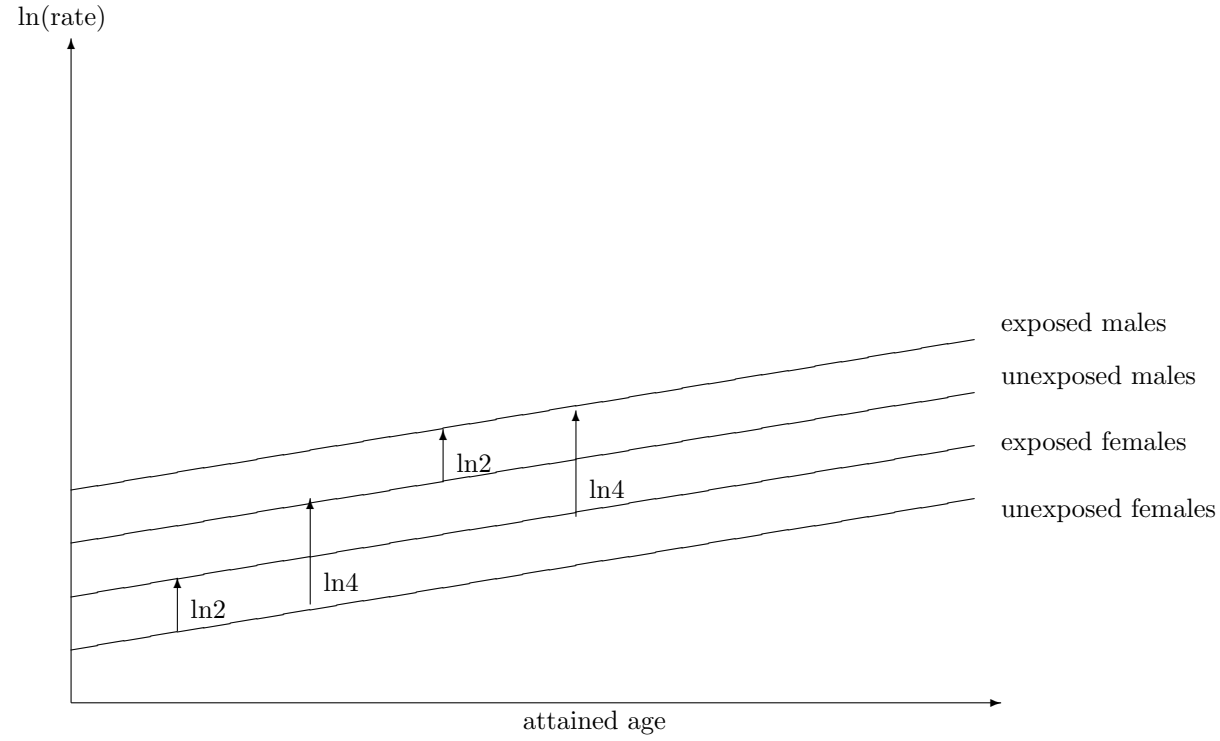
The critical value of a χ^2 with 4 degrees of freedom is 13.3 at the 1 % significance level. Since our test statistic of $26.8 > 13.3$ the LR test is highly significant and we can reject the null hypothesis that all interaction effects are 0.

This test would of course also had been significant if we had set the significance level to 95 % (in this case the critical value of the χ^2_4 distribution would have been 9.49).

2. (a) (4 marks) It is known that the incidence rate of a certain disease depends on attained age and gender. A cohort study is conducted to determine whether the incidence rate depends on a binary exposure of interest. A Cox regression model is fitted to the data with attained age as the timescale. The estimated hazard ratio for gender (males/females) was 4 and the estimated hazard ratio for exposure (exposed/unexposed) was 2. Assume that there was no evidence of interaction between any of the variables.

Imagine that, for unexposed females, the association between the natural logarithm of the incidence rate and attained age has the form shown in the figure below. Assuming that the assumptions of the Cox model are appropriate, complete the figure below by drawing lines for the other 3 combinations of gender and exposure. The aim of this exercise is for you to demonstrate that you understand the assumptions of the Cox model. You should indicate how the estimated hazard ratios are represented on the graph.

Answer:



(b) (4 marks) Now imagine that we split attained age into three categories and fit a Poisson regression model (with attained age, exposure, and gender as explanatory variables) to these same data. On the graph below, plot the fitted (i.e. predicted by the model) value of the natural logarithm of the incidence rate as a function of attained age for each of the four combinations of gender and exposure. The aim of this question is for you to demonstrate that you understand the fundamental difference between Cox regression and Poisson regression. You may assume that the estimated incidence rate ratio for gender (males/females) was 4 and the estimated incidence rate ratio for exposure (exposed/unexposed) was 2. You should indicate how these estimates are represented on the graph.

Answer: Below is one example where the underlying time scale has been categorised into 3 timebands.