Solutions, Biostat III exam, Nov 23, 2012

1 Section 1

1. (a) Inmates who received financial aid after being released from prison had an approximately 29% lower rate of criminal recidivism compared to those who received no financial aid. This incidence rate ratio is adjusted for age at release from prison, previous full-time work experience as well as the underlying time-scale (i.e., weeks since release from prison). The effect of financial aid on criminal recidivism is not statistically significant (p = 0.077).

   (b) The null hypothesis to be tested is:

   $$H_0 : \text{The parameter representing the effect of age (i.e., } \ln(HR_{\text{age}}) \text{ or } \beta_{\text{age}}) = 0$$

   against

   $$H_A : \text{The parameter representing the effect of age (i.e., } \ln(HR_{\text{age}}) \text{ or } \beta_{\text{age}}) \neq 0$$

   Under the null hypothesis, the test statistic (Z) is assumed to follow a standard normal distribution.

   (c)

   $$HR = 0.944855^{40-35} = 0.944855^5 \approx 0.75$$

   (d) No, I do not agree with my colleague. It is correct that there may very well be an association between social class and recidivism. However, the fact that social class might be an independent risk factor does not necessarily mean that social class is a confounder. In this study financial aid was randomised to half of the inmates in the study (this is written in the description of the data). This suggests that there is no association between social class and financial aid so there is no need to adjust for social class for the purposes of this study.

2. (a) Among inmates with no previous full-time work experience the effect of receiving financial aid is 0.698. Among inmates with full-time work experience the corresponding HR $= 0.698 \times 1.051 \approx 0.73$.

   (b) In order to answer this question we can conduct a likelihood ratio test where we compare Model A (main effects model) to Model B (model that also contains an interaction term for the effects of financial aid and previous work experience).

   The null hypothesis to be tested is:

   $$H_0 : \text{The parameter representing the interaction effect is 0.}$$

   against

   $$H_A : \text{The parameter representing the interaction effect is not 0.}$$

   The test statistic is defined as

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

   and has a χ^2_p distribution under H_0, where p is equal to the degrees of freedom. Note that the degrees of freedom is determined by how many 'extra' parameters 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 A (log likelihood: -664.94013) and model B (log likelihood: -664.93151) and get -0.00862. This is multiplied by -2 to get 0.01724 which is the value of our test statistic.

The critical value of a χ^2 with 1 degrees of freedom is 3.841 at the 95% significance level. Since our test statistic of 0.01724< 3.841 the LR test is not statistically significant and we cannot reject the null hypothesis that the interaction effect is 0. In other words there is no formal evidence that the effect of financial aid is modified by prior work experience.

2 Section 2

3. (a)

$$S(12) = S(11)(1 - \frac{D}{\text{number at risk}}) = (0.9606)(1 - \frac{2}{415}) \approx 0.9560$$

(b) 12 weeks after release from prison 95.6% of all inmates had not yet re-offended.

(c) A hazard rate attempts to estimate the instantaneous risk of experiencing the outcome of interest. In other words, at each given time point during follow up the hazard rate provides an estimate of the risk of experiencing the event of interest in the very next instant. [To get full marks you needed to comment that the hazard is an instantaneous measure and that it is a measure of risk.]

(d) The rate of criminal recidivism seems to increase with elapsed time since release from prison.

[Additional comments that were not required to obtain full marks]

It appears as if the incidence rate peaks around 40 weeks, however the dip after 40 weeks may very well just reflect random variation in the data or an artifact of the smoothing procedure. To formally characterize the association between criminal recidivism and time since release from prison we need to use a more sophisticated statistical method (such as a regression model).

4. (a) We first calculate the HR and a 95% confidence interval based on the provided Stata output.

$$HR = e^{-.5824554} \approx 0.56$$
$$LCL = e^{-.5824554 - 1.96 \times 0.1881361} \approx 0.38$$
$$UCL = e^{-.5824554 + 1.96 \times 0.1881361} \approx 0.81$$

Adjusted for elapsed time since release from prison, inmates with previous full-time work experience had approximately 44% reduced rate of criminal recidivism compared to inmates with no previous full-time work experience. The 95% confidence interval for the estimated HR ranges from 0.38 to 0.81. Since the confidence interval does not include 1.00 the reported HR is statistically significant.

(b) Under the proportional hazards assumption we would expect to see that the slope of the two KM-curves would be roughly the same. However, it is difficult to assess the proportional hazards assumption in a KM graph (unless there are strong departures from this assumption). In the graph of the hazards we would look for something that resembles parallelism since the y-axis is on a log scale.

(c) We could study the scaled Schoenfeld residuals and formally conduct a test if there is evidence of a linear trend through the residuals (Grambsch-Thernau test). Alternatively, we could split the time scale into time bands and fit a Cox model that includes interaction terms between the time bands and the variable whose effect we suspect is non-proportional. We can then perform, for example, a likelihood ratio test to check if the interaction is statistically significant.

5. (a)

$$HR_{regional/localised} = \frac{D_{regional}/p - time_{regional}}{D_{distant}/p - time_{distant}} = \frac{218/1.5002}{1013/38.6266} = 5.54$$

(b)

$$HR_{regional/localised} = \frac{5.116341}{1.018815} = 5.02$$

The estimated HR from the Cox model is different from the HR that we estimated in part a). The reason for the difference is that we adjust for the time scale (time since diagnosis) in the Cox model. In part a) we implicitly assumed that the rates in the two groups were constant across the entire time scale.

(c) We could replicate the results from the Cox model if we would fine split follow-up (i.e., split the data at each event time) and include a variable representing the time bands in our Poisson model. The variable representing time would have to be modelled categorically or non-linearly using a spline function.

6. (a)

$$\ln(\text{rate}_{\text{males}}) = \beta_0 = -3.688015$$

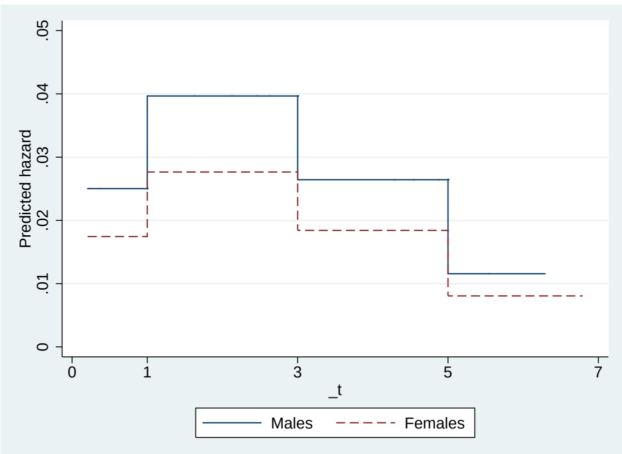
(b)

$$\ln(\text{rate}_{\text{females}}) = \beta_0 + \beta_{\text{females}} = -3.688015 - 0.3609177$$

(c)

$$\ln(\text{rate}_{\text{females}}) - \ln(\text{rate}_{\text{males}}) = \ln(HR)$$

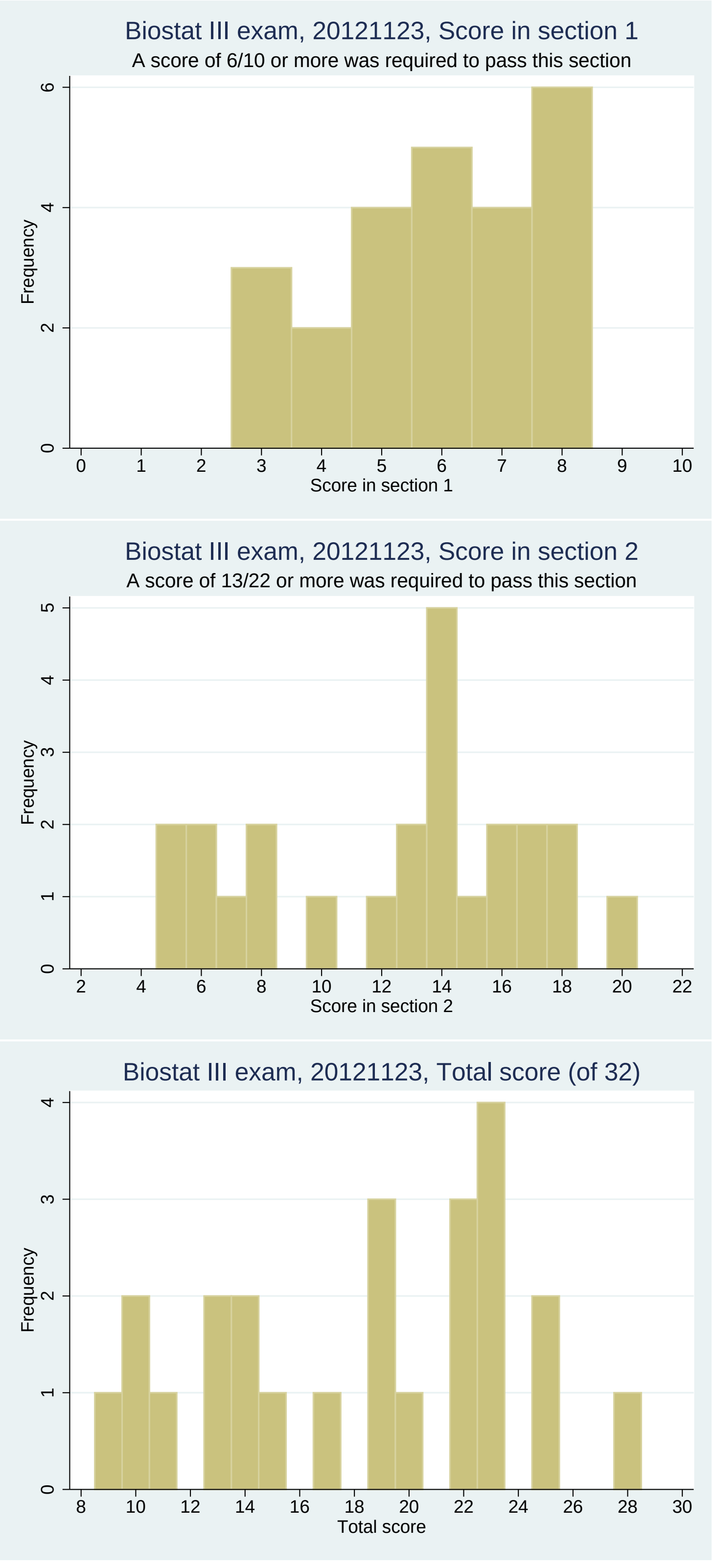
(d)



[In general marks were awarded as follows: 1 point for getting the cutpoints correct, 1 point for correct calculations, and 1 point for demostrating piecewise linearity]

(e) Yes

3 Results


Biostat III exam, 20121123, Score in section 1
A score of 6/10 or more was required to pass this section
Frequency
0
2
4
6
0
1
2
3
4
5
6
7
8
9
10
Score in section 1
Biostat III exam, 20121123, Score in section 2
A score of 13/22 or more was required to pass this section
Frequency
0
1
2
3
4
5
2
4
6
8
10
12
14
16
18
20
22
Score in section 2
Biostat III exam, 20121123, Total score (of 32)
Frequency
0
1
2
3
4
8
10
12
14
16
18
20
22
24
26
28
30
Total score