BIOSTAT III: Survival Analysis

Examination

February 14, 2014

Time: 9:00–11.30

Exam room location: Jacob Berzelius, Berzelius väg 3, Karolinska Institutet

Code (please do not write your name):

- Time allowed is 2 1/2 hours.
- Please try and write your answers on the exam sheet. You may use separate paper if absolutely necessary. Your working and motivation for your answer, not just the final answer, will be assessed when grading the examination.
- The exam contains 2 sections; the first section tests your knowledge in general epidemiological concepts in a survival analysis framework whereas the second section focusses on more specific topics in survival analysis. Each section contains multiple questions (with several parts). The marks available for each part are indicated.
- A score of 6 marks or more out of 12 in the first section, and a score of 11 or more out of 22 in the second section will be required to obtain a passing grade.
- The questions may be answered in English or Swedish (or a combination thereof).
- A non-programmable scientific calculator (i.e., with ln() and exp() functions) may be useful. You may not use a mobile phone or other communication device as a calculator or for any other purpose.
- The exam is not 'open book' but each student will be allowed to bring one A4 sheet of paper into the exam room which may contain, for example, hand-written notes or photocopies from textbooks/lecture notes etc. Both sides of the page may be used.
- The exam supervisors have been advised not to answer any questions you may have regarding the content of the exam. If you believe a question contains an error or is ambiguous then please write a note with your answer indicating how you have interpreted the question.
- Tables of critical values of the χ^2 distribution are provided on the last page.

Section 1

The following questions test your knowledge of general concepts in statistical modelling of epidemiological data.

The Montana study of smelter workers (Sv: smältverksarbetare)

For the first two questions in this exam, we use data from Breslow and Day (1987) on the Montana study of smelter workers. The cohort was defined by company employment records for those employed for at least one year before 1 December 1957. Environmental measures taken in the smelter enabled each work area to be categorized on a 1 to 10 scale for arsenic exposure. Each worker's arsenic exposure was determined using the time and place of employment for each job held within the smelter. The arsenic exposure variable represents the number of years that a worker was exposed to moderate or high arsenic exposure. The main aim of this analysis is to assess whether arsenic exposure is associated with respiratory cancer death.

The following Stata output shows the codebook for the variables used in the analyses for this exam. The two time scales of interest were attained age and attained calendar period. The person-time has already been split by these two time scales. Workers were grouped by arsenic exposure and period of hire. The data were aggregated by attained age, attained calendar period, period of hire and arsenic exposure, with variables for person-time of observation and the number of respiratory cancer deaths.

```
-------------------------------------------------------------------------------
age_group                                          Attained age group (years)
-------------------------------------------------------------------------------

                 type:  numeric (float)
                label:  age_group

                range:  [1,4]                        units:  1
        unique values:  4                    missing .:  0/114

           tabulation:  Freq.   Numeric  Label
                           24         1  40-49
                           28         2  50-59
                           31         3  60-69
                           31         4  70-79

-------------------------------------------------------------------------------
calendar_period                                      Attained calendar period
-------------------------------------------------------------------------------

                 type:  numeric (float)
                label:  calendar_period

                range:  [1,4]                        units:  1
        unique values:  4                    missing .:  0/114

           tabulation:  Freq.   Numeric  Label
                           30         1  1938-1949
                           32         2  1950-1959
                           28         3  1960-1969
                           24         4  1970-1977
```

```
-------------------------------------------------------------------------------
arsenic                      Number of years of moderate to high exposure to arsenic
-------------------------------------------------------------------------------

                 type:  numeric (float)
                label:  arsenic

                range:  [0,1]                        units:  1
        unique values:  2                    missing .:  0/114

           tabulation:  Freq.   Numeric  Label
                           29         0  0-0.9
                           85         1  1.0+

-------------------------------------------------------------------------------
period_of_hire                           Calendar period when the subject was hired
-------------------------------------------------------------------------------

                 type:  numeric (float)
                label:  period_of_hire

                range:  [1,2]                        units:  1
        unique values:  2                    missing .:  0/114

           tabulation:  Freq.   Numeric  Label
                           52         1  <1925
                           62         2  1925+

-------------------------------------------------------------------------------
resp_ca_deaths                                 Number of respiratory cancer deaths
-------------------------------------------------------------------------------

                 type:  numeric (float)

                range:  [0,19]                       units:  1
        unique values:  15                   missing .:  0/114

                 mean:   2.42105
            std. dev:   3.30163

          percentiles:        10%       25%       50%       75%       90%
                               0         0         1         3         7

-------------------------------------------------------------------------------
person_years                                          Person-years of observation
-------------------------------------------------------------------------------

                 type:  numeric (float)

                range:  [4.25,12451.29]              units:  .01
        unique values:  114                  missing .:  0/114

                 mean:   1096.41
            std. dev:    2123.1

          percentiles:        10%       25%       50%       75%       90%
                            54.74    127.51   335.165    933.45   2511.97
```

Section 2

The following questions test your knowledge of concepts that are of special interest in survival analysis. For this section, we use the following dataset. The Radiation Therapy Oncology Group (RTOG) carried out a *randomised controlled trial* for patients with cancer of the oropharanynx (Kalbfleisch and Prentice 2002, Dataset II). Patients were randomly assigned to one of two trial arms, including (i) *standard therapy*, being radiation therapy alone, and (ii) *test therapy*, being radiation therapy with a chemotherapeutic agent. The primary outcome was death due to any cause. In the following, T staging describes the size of the original (primary) tumour and whether it has invaded nearby tissue. The variables are described below.

```
. codebook Status

-------------------------------------------------------------------------------
Status                                                          Event indicator
-------------------------------------------------------------------------------

                  type:  numeric (float)
                 label:  Death

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  0/195

            tabulation:  Freq.   Numeric  Label
                            53         0  Censored
                           142         1  Death

. codebook Time

-------------------------------------------------------------------------------
Time                                                Days from diagnosis (days)
-------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [11,1823]                    units:  1
         unique values:  177                      missing .:  0/195

                  mean:   558.728
             std. dev:   418.718

           percentiles:        10%       25%       50%       75%       90%
                               134       238       445       782      1250


. codebook TrtGp

-------------------------------------------------------------------------------
TrtGp                                                            Treatment arm
-------------------------------------------------------------------------------

                  type:  numeric (float)
                 label:  TrtGp

                 range:  [1,2]                        units:  1
         unique values:  2                        missing .:  0/195

            tabulation:  Freq.   Numeric  Label
                           100         1  Standard
                            95         2  Test
```

```
. codebook AgeGp

-------------------------------------------------------------------------------
AgeGp                                                        Age group (years)
-------------------------------------------------------------------------------

                 type:  numeric (float)
                label:  AgeGp

                range:  [0,3]                        units:  1
        unique values:  4                    missing .:  0/195

           tabulation:  Freq.   Numeric  Label
                           35         0  20-
                           57         1  50-
                           64         2  60-
                           39         3  70-

. codebook Tstaging

-------------------------------------------------------------------------------
Tstaging                                                             T staging
-------------------------------------------------------------------------------

                 type:  numeric (float)
                label:  Tstaging2

                range:  [1,3]                        units:  1
        unique values:  3                    missing .:  0/195

           tabulation:  Freq.   Numeric  Label
                           35         1  primary, <=4cm
                           93         2  primary, >4cm
                           67         3  massive invasive tumor

. codebook Case

-------------------------------------------------------------------------------
Case                                                                 Case ID
-------------------------------------------------------------------------------

                 type:  numeric (float)

                range:  [1,195]                      units:  1
        unique values:  195                  missing .:  0/195

                 mean:        98
            std. dev:   56.4358

          percentiles:       10%       25%       50%       75%       90%
                              20        49        98       147       176
```

Table A3 Critical Values of Chi-Square

df	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
1	2.706	3.841	6.635
2	4.605	5.991	9.210
3	6.251	7.815	11.345
4	7.779	9.488	13.277
5	9.236	11.070	15.086
6	10.645	12.592	16.812
7	12.017	14.067	18.475
8	13.362	15.507	20.090
9	14.684	16.919	21.666
10	15.987	18.307	23.209
11	17.275	19.675	24.725
12	18.549	21.026	26.217
13	19.812	22.362	27.688
14	21.064	23.685	29.141
15	22.307	24.996	30.578
16	23.542	26.296	32.000
17	24.769	27.587	33.409
18	25.989	28.869	34.805
19	27.204	30.144	36.191
20	28.412	31.410	37.566
21	29.615	32.671	38.932
22	30.813	33.924	40.289
23	32.007	35.172	41.638
24	33.196	36.415	42.980
25	34.382	37.652	44.314
30	40.256	43.773	50.892
35	46.059	49.802	57.342
40	51.805	55.758	63.691
45	57.505	61.656	69.957
50	63.167	67.505	76.154
60	74.397	79.082	88.379
70	85.527	90.531	100.425
80	96.578	101.879	112.329
90	107.565	113.145	124.116
100	118.498	124.432	135.807

The value tabulated is c such that $P(\chi^2 \geq c) = \alpha$.