BIOSTAT III: Survival Analysis

Examination

November 23, 2012

Time: 9:00–11.30

Exam room location: Lecture hall MTC,
Nobels väg 16, 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.
- 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) will most probably 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.

Description of the data sets used in this exam

The recidivism data

For the first four questions of this exam we have used data from a study by Rossi, Berk, and Lenihan (1980) on recidivism (i.e., reoffending) of 432 prisoners during the first year after their release from Maryland state prisons. The aim of the research was to determine the efficacy of financial aid to released inmates as a means of reducing recidivism. Half of the inmates were randomly assigned to financial aid. They were followed for one year after their release and were interviewed monthly during that period. Data on arrests were taken from police and court records.

The following Stata output shows output from the `stset` command and frequency tables for some of the variables used in the analyses for this exam.

```
. /** stset the data using time since release from prison as the timescale
(in complete weeks) **/

. stset week, failure(arrest)

     failure event:  arrest != 0 & arrest < .
obs. time interval:  (0, week]
 exit on or before:  failure

------------------------------------------------------------------------------
      432  total obs.
        0  exclusions
------------------------------------------------------------------------------
      432  obs. remaining, representing
      114  failures in single record/single failure data
    20127  total analysis time at risk, at risk from t =         0
                              earliest observed entry t =         0
                                   last observed exit t =        52



------------------------------------------------------------------------------
fin                             The inmate received financial aid after release
------------------------------------------------------------------------------

                  type:  numeric (double)
                 label:  fin_lab

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  0/432

            tabulation:  Freq.   Numeric  Label
                           216         0  No financial aid
                           216         1  Financial aid


------------------------------------------------------------------------------
wexp             The inmate had full-time work experience before incarceration
------------------------------------------------------------------------------

                  type:  numeric (double)
                 label:  wexp

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  0/432

            tabulation:  Freq.   Numeric  Label
                           185         0  No
                           247         1  Yes
```

```
-------------------------------------------------------------------------------
age                             Age in years at the time of release from prison
-------------------------------------------------------------------------------

                  type:  numeric (double)

                 range:  [17,44]                     units:  1
         unique values:  28                       missing .:  0/432

                  mean:   24.5972
             std. dev:   6.11338


-------------------------------------------------------------------------------
educ                                    Highest level of completed schooling
-------------------------------------------------------------------------------

                  type:  numeric (double)
                 label:  educ

                 range:  [2,6]                       units:  1
         unique values:  5                        missing .:  0/432

            tabulation:  Freq.   Numeric  Label
                            24         2  6th grade or less
                           239         3  7th to 9th grade
                           119         4  10th to 11th grade
                            39         5  12th grade
                            11         6  some college
```

The melanoma data

For questions five and six in this exam we analyse melanoma data from Finland. The aim is to study cause-specific survival from melanoma with respect to patient and disease characteristics such as age at diagnosis, year of diagnosis, sex and stage at diagnosis. The underlying time scale for the analysis is time since diagnosis.

The following Stata output shows output from the `stset` command and frequency tables for some of the variables used in the analysis.

```
. stset surv_mm, failure (status == 1) id(id) scale(12)

                id:  id
     failure event:  status == 1
obs. time interval:  (surv_mm[_n-1], surv_mm]
 exit on or before:  failure
    t for analysis:  time/12

--------------------------------------------------------------------------------
     7775  total obs.
        0  exclusions
--------------------------------------------------------------------------------
     7775  obs. remaining, representing
     7775  subjects
     1913  failures in single failure-per-subject data
 51269.71  total analysis time at risk, at risk from t =         0
                              earliest observed entry t =         0
                                   last observed exit t =  20.95833




--------------------------------------------------------------------------------
agegrp                                          Age in 4 categories
--------------------------------------------------------------------------------

                  type:  numeric (byte)
                 label:  agegrp

                 range:  [0,3]                        units:  1
         unique values:  4                        missing .:  0/7775

            tabulation:  Freq.   Numeric  Label
                          2046         0  0-44
                          2238         1  45-59
                          2280         2  60-74
                          1211         3  75+

--------------------------------------------------------------------------------
year8594                                      Year of diagnosis 1985-94
--------------------------------------------------------------------------------

                  type:  numeric (byte)
                 label:  year8594

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  0/7775

            tabulation:  Freq.   Numeric  Label
                          3031         0  Diagnosed 75-84
                          4744         1  Diagnosed 85-94
```

```
------------------------------------------------------------------------------
stage                                   Clinical stage at diagnosis
------------------------------------------------------------------------------

                 type:  numeric (byte)
                label:  stage

                range:  [0,3]                        units:  1
        unique values:  4                    missing .:  0/7775

           tabulation:  Freq.   Numeric  Label
                         1631         0  Unknown
                         5318         1  Localised
                          350         2  Regional
                          476         3  Distant

------------------------------------------------------------------------------
sex                                                                       Sex
------------------------------------------------------------------------------

                 type:  numeric (byte)
                label:  sex

                range:  [1,2]                        units:  1
        unique values:  2                    missing .:  0/7775

           tabulation:  Freq.   Numeric  Label
                         3680         1  Male
                         4095         2  Female
```

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 $P(\chi^2 \geq c) = \alpha$.