Australian National University

RESEARCH SCHOOL OF FINANCE, ACTUARIAL STUDIES AND APPLIED STATISTICS

First Semester Mid-Semester Examination (2014)

Survival Models / Biostatistics (STAT3032/7042/8003)

Writing period:	*1 hour duration*
Study period:	*15 minutes duration*
Permitted materials:	*Non-programmable calculator, one page of A4 size paper, dictionary*
Total marks:	*30 (undergraduates) / 35 (postgraduates) marks*

INSTRUCTIONS TO CANDIDATES:

- *Postgraduates should attempt all questions. Undergraduates should only attempt questions 1 to 4.*
- *To ensure full marks show all the steps in working out your solutions. Marks may be deducted for failure to show appropriate calculations or formulae.*
- *All questions are to be completed in the script book provided.*
- *All answers should be rounded to 4 decimal places.*

Question 1 [5 marks]

Given that $p_x = 0.99$, $p_{x+1} = 0.985$, ${}_3p_{x+1} = 0.95$ and $q_{x+3} = 0.02$.

(a) **[2 marks]** Calculate ${}_2p_x$.

(b) **[3 marks]** Calculate ${}_3p_x$.

Solution:

$$\begin{aligned} p_{x+3} &= 1 - q_{x+3} = 0.98 \\ {}_2p_x &= p_x p_{x+1} = 0.9752 \\ {}_2p_{x+1} &= {}_3p_{x+1}/p_{x+3} = 0.9694 \\ {}_3p_x &= p_{x\,2}p_{x+1} = 0.9597 \end{aligned}$$

Question 2 [10 marks]

(For each part, you will gain 2 marks for a correct answer, be penalized 2 marks for an incorrect answer, and score 0 if no answer is given.) Answer each question "TRUE" or "FALSE". In each case, write the whole word. It is **not** acceptable to write only "T" or "F" and answers presented in this form **will be graded incorrect**

(a) **[2 marks]** For a constant force of mortality μ, $m_x = \mu$.
[**Hint**: $m_x = \frac{d_x}{\int_0^1 l_{x+t}dt} = \frac{q_x}{\int_0^1 {}_tp_x dt}$]

(b) **[2 marks]** Parametric estimates are always more efficient than non-parametric estimates.

(c) **[2 marks]** If you know the hazard function, then you can determine the corresponding survival function, and vice versa.

(d) **[2 marks]** For curtate expectation of time, $e_x = {}_np_x(n + e_{x+n})$ for all $0 < n < x$.

(e) **[2 marks]** The survival functions for different covariate values cannot cross for both Cox regression and KM estimator.

Solution:
TRUE FALSE TRUE FALSE FALSE

Question 3 [10 marks]

Data are available from a small study on claim incidence in PHI. A subset of policyholders all aged 50 with no previous claims history is monitored. Policyholders are classified by sex, (Male, Female). The data, times to claim (in months), are given in the table below; the * indicates that an observation was censored.

Male: 2* 3 3 8* 12* 14 16* 21*
Female: 3* 5 7 9* 13 17* 24*

(a) [**5 marks**] Find the Kaplan-Meier estimate of the survivor function for all policy-holders combined.

Solution:

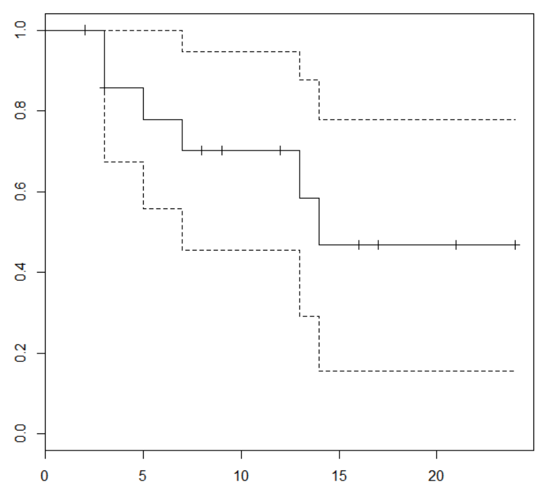
```
> summary(survfit(formula = Surv(times, ind)~ 1, conf.type ="plain"))

Call: survfit(formula = Surv(times, ind) ~ 1, conf.type = "plain")

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
    3     14       2    0.857  0.0935        0.674        1.000
    5     11       1    0.779  0.1129        0.558        1.000
    7     10       1    0.701  0.1257        0.455        0.948
   13      6       1    0.584  0.1495        0.291        0.877
   14      5       1    0.468  0.1589        0.156        0.779
```

(b) [**3 marks**] Roughly plot your estimate of the survivor function.

Solution:



(c) [**2 marks**] Estimate $S(8)$ and explain briefly why the estimates of $S(8)$ and $S(9)$ are the same for this example.

Solution:
$S(8)$ is 0.701, the same as $S(7)$ and $S(9)$. The reason is that there was no claim

between 8th and 9th months.

Question 4 [5 marks]

The following force of mortality is assumed to hold for an individual aged x :

$$\mu_x(t) = \frac{1}{10-t}, \ 0 \le t < 10$$

(a) **[3 marks]** Assuming this force of mortality holds, calculate the complete expected lifetime e_x^0

Solution:

$$\begin{aligned}
{}_tp_x &= \exp\{-\int_0^t (10-s)^{-1}ds\} \\
&= \exp\{-[-\log(10-s)]_0^t\} \\
&= \exp\{-[-\log(10-t)+\log(10)]\} \\
&= \exp\{\log(10-t)-\log(10)\} \\
&= \frac{10-t}{10} \\
&= 1-\frac{t}{10} \\
e_x^0 &= \int_0^{10}(1-\frac{t}{10})dt \\
&= [t-\frac{t^2}{20}]_0^{10} \\
&= 5
\end{aligned}$$

(b) **[2 marks]** In lecture, we show very briefly that $e_x^0 \le n + e_{x+n}^0$ with some "naive" assumptions. Use $e_x^0 = \int_0^\infty {}_tp_x dt$ to show why it holds for the general case. [**Hint**: $e_x^0 = \int_0^n {}_tp_x dt + \int_n^\infty {}_tp_x dt$ and ${}_tp_x \le 1$]

Solution:

$$\begin{aligned}
e_x^0 &= \int_0^n {}_tp_x dt + \int_n^\infty {}_tp_x dt \\
&\le n + \int_n^\infty {}_np_{x\,t-n}p_{x+n} dt \\
&\le n + \int_n^\infty {}_{t-n}p_{x+n} dt
\end{aligned}$$

$$= n + \int_0^\infty {}_sp_{x+n} ds$$
$$= n + e^0_{x+n}$$

Question 5 [5 marks] (For students enrolled in STAT7042/8003 ONLY)

Suppose $E[Y_1] = 1$, $E[Y_2] = 2$, $E[Y_3] = 3$, $Var[Y_1] = 1$, $Var[Y_2] = 2$, $Var[Y_3] = 3$, $Cov(Y_1, Y_2) = Cov(Y_1, Y_3) = 0$ and $Cov(Y_2, Y_3) = 1$. For a variable $W = \frac{Y_3}{Y_1+Y_2}$, use the delta method to find:

(a) [**1 mark**] The approximate mean of W.

Solution:
$E(W) = \frac{E(Y_3)}{E(Y_1)+E(Y_2)} = 3/(1+2) = 1$

(b) [**4 marks**] The approximate variance of W.

Solution:

$$E[(\partial W/\partial Y)^T] = \begin{pmatrix} -\frac{E(Y_3)}{(E(Y_1)+E(Y_2))^2} & -\frac{E(Y_3)}{(E(Y_1)+E(Y_2))^2} & \frac{1}{E(Y_1)+E(Y_2)} \end{pmatrix}$$
$$= \begin{pmatrix} -1/3 & -1/3 & 1/3 \end{pmatrix}$$
$$\sum = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 3 \end{pmatrix}$$
$$Var(W) = E[(\partial W/\partial Y)^T] \sum E(\partial W/\partial Y)$$
$$= \begin{pmatrix} -1/3 & -1/3 & 1/3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 3 \end{pmatrix} \begin{pmatrix} -1/3 \\ -1/3 \\ 1/3 \end{pmatrix}$$
$$= 4/9$$

END OF EXAMINATION