Australian National University

**RESEARCH SCHOOL OF FINANCE, ACTUARIAL STUDIES AND STATISTICS**

*First Semester Mid-Semester Examination (2018)*

# Survival Models/Biostatistics (STAT 3032/4072/7042/8003)

| | |
|---|---|
| *Writing period:* | *1.5 hours duration* |
| *Study period:* | *15 minutes duration* |
| *Permitted materials:* | *Non-programmable calculator, dictionary, one A4 sized sheet of paper with notes on both sides* |
| *Total marks:* | *50 marks* |

*INSTRUCTIONS TO CANDIDATES:*

- *Students should attempt all questions.*
- *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 [4 marks]

In 1729 de Moivre hypothesized the following force of mortality for an individual at age $x$ (e.g. $x$ is the future lifetime of an individual aged 0):

$$\mu_x = (m - x)^{-1}, \quad 0 \le x < m.$$

Assuming this force of mortality holds,

(a) [**3 marks**] Calculate $S(x)$ the probability that an individual aged 0 survives to age $x$.

> **Solution:**
>
> $$\begin{aligned} S(x) &= \exp\left\{-\int_0^x (m-t)^{-1} dt\right\} \\ &= \exp\left\{-\left[-\log(m-t)\right]_0^x\right\} \\ &= \exp\left\{\log(m-x) - \log(m)\right\} \\ &= 1 - \frac{x}{m} \end{aligned}$$

(b) [**1 mark**] Further explain in de Moivres law, why is $x$ restricted to be in the range $0 \le x < m$.

> **Solution:**
> $x$ is restricted to be in the range $0 \le x < m$ so that the force of mortality remains positive and defined, or $S(x)$ ranges from 0 to 1.

## 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**] ${}_6p_{34} \cdot p_{33} \cdot (1 - q_{32})$ is equal to ${}_7p_{32}$.

(b) [**2 marks**] If the force of mortality function $\mu_x$ is assumed to follow Makehams law, this means that for each one year increment in age, $\mu_x$ increases by a constant scale.

(c) [**2 marks**] Gompertz Law is appropriate for modelling the force of mortality for humans over the age range 0 to 40 years.

(d) [**2 marks**] Treating censored observations as times of death can result in underestimating the survival function (e.g. $\hat{S}(t)$ is smaller than the true value $S(t)$ for some $t$).

(e) **[2 marks]** The complete expected future lifetime $e_x^0$ must be greater than or equal to the curtate expectation of life $e_x$.

**Solution:**

FALSE FALSE FALSE TRUE TRUE

# Question 3 [7 marks]

For a particular population it is shown that $l_x = 50 - 0.5x$, $0 \leq x \leq 100$. Using this information about the number of lives aged $x$ exact, calculate the following:

(a) **[2 marks]** the force of mortality at age 30.

**Solution:**

$$\mu_x = -\frac{1}{l_x}\frac{dl_x}{dx} = \frac{1}{100 - x}$$
$$\mu_{20} = \frac{1}{100 - 30} = 0.0143$$

(b) **[2 marks]** the complete expectation of life at age 30.

**Solution:**

$$e_{30}^0 = \int_0^{100-30} {}_tp_{30} dt = \int_0^{70} (1 - \frac{t}{70})dt = 35$$

(c) **[3 marks]** the average age of individuals who die between ages 60 and 65.

**Solution:**

$$60 + \int_0^5 \frac{t l_{60+t}\mu_{60+t}}{l_{60} - l_{65}} dt = 60 + \int_0^5 \frac{t(50 - 30 - 0.5t)(1/(100 - 60 - t))}{2.5} = 62.5$$

# Question 4 [12 marks]

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

2, 3*, 3*, 8, 12, 14*, 16, 21

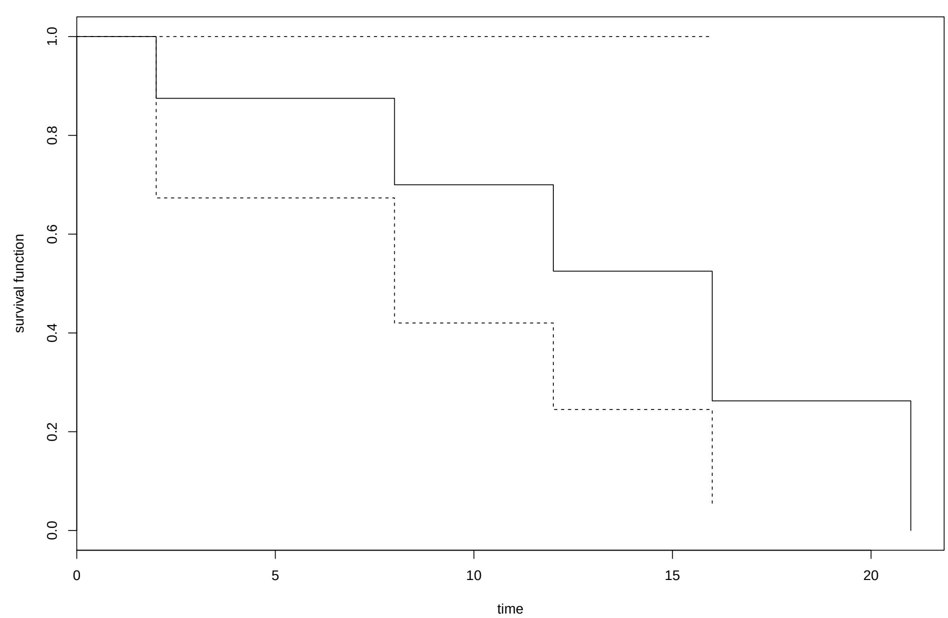
(a) [**5 marks**] Calculate the Kaplan-Meier estimate of the survivor function $S(t)$ for these policyholders. You should also provide standard errors for your estimated function.

**Solution:**

| time | n.risk | n.event | survival | std.err | lower 95% CI | upper 95% CI |
|---|---|---|---|---|---|---|
| 2 | 8 | 1 | 0.875 | 0.117 | 0.673 | 1 |
| 8 | 5 | 1 | 0.700 | 0.182 | 0.420 | 1 |
| 12 | 4 | 1 | 0.525 | 0.204 | 0.245 | 1 |
| 16 | 2 | 1 | 0.263 | 0.212 | 0.054 | 1 |
| 21 | 1 | 1 | 0.000 | NaN | NA | NA |

(b) [**2 marks**] Roughly plot your estimates of the survivor function, you should label all the survival functions and times at death.

**Solution:**




(c) [**2 marks**] Estimate $S(4)$ and explain why the estimates of $S(4)$ and $S(5)$ are the same.

**Solution:**
$S(4) = 0.875$. Nobody claimed between 4 and 5 months so the estimate of $S(t)$ does not change.

(d) [**3 marks**] Provide an estimate of the mean time to claim for policyholders.

**Solution:**
One way to answer this question is to approximate the integral $\int_0^\infty {}_tp_x dt$. This can be done by computing the area under the estimated KM survival function. This area is equal to

$$(2-0)\times 1 + (8-2)\times 0.875 + (12-8)\times 0.700$$
$$+ (16-12)\times 0.525 + (21-16)\times 0.263 = 13.465$$

# Question 5 [7 marks]

The Uniform Distribution of Deaths (UDD) assumes that the pdf of lifetime $T_x$ follows a uniform distribution for $0 < t < 1$.

(a) [**2 marks**] Show that UDD implies that ${}_sq_x = s \cdot q_x$, where $0 < s < 1$.

**Solution:**
Since $\mu_x(t) = f_x(t)/S_x(t) = M/S_x(t)$, then ${}_tp_x\mu_{x+t} = M$.
Then ${}_sq_x = \int_0^s {}_sp_x\mu_{x+s}ds = sM$ and $q_x = \int_0^1 {}_sp_x\mu_{x+s}ds = M$. So ${}_sq_x = s \cdot q_x$.

(b) [**2 marks**] Demonstrate that for any $0 \le a < b$, ${}_{b-a}q_{x+a} = 1 - \frac{{}_bp_x}{{}_ap_x}$.

**Solution:**
${}_ap_x -_b p_x =_a p_x -_{b-a+a} p_x =_a p_x -_a p_x \cdot {}_{b-a}p_{x+a} =_a p_x -_a p_x(1 -_{b-a} q_{x+a})$
Then ${}_ap_x -_b p_x =_a p_x \cdot_{b-a} q_{x+a}$ and ${}_{b-a}q_{x+a} = 1 - \frac{{}_bp_x}{{}_ap_x}$.

(c) [**3 marks**] Prove that for $0 \le a < b \le 1$, we have ${}_{b-a}q_{x+a} = \frac{(b-a)q_x}{1 - a \cdot q_x}$ under UDD and use it to calculate ${}_{0.1}q_{x+0.5}$, given that ${}_{0.8}q_x = 0.2$.

**Solution:**
$${}_{b-a}q_{x+a} = 1 - \frac{{}_bp_x}{{}_ap_x} = \frac{{}_ap_x -_b p_x}{1 -_a q_x} = \frac{{}_bq_x -_a q_x}{1 -_a q_x} = \frac{(b-a)q_x}{1 - a \cdot q_x}.$$
Since ${}_{0.8}q_x = 0.2$, we know that $q_x = 0.2/0.8 = 0.25$ and ${}_{0.1}q_{x+0.5} = \frac{(0.1)q_x}{1 - 0.5 \cdot q_x} =$ 0.0286.

## Question 6 [10 marks]

The lifetimes of a certain species of insect, denoted $x$, are believed to follow a Pareto distribution. The density of the Pareto distribution is given by:

$$\frac{\theta\lambda^\theta}{x^{\theta+1}}, \theta > 0, \lambda > 0, x \geq \lambda.$$

A sample of six insects had the following survival times: 4, 5.5, 6.5, 7, 8, 11. It is known that $\lambda = 1$ for this species of insect.

(a) [**3 marks**] Compute the maximum likelihood estimate of $\theta$

**Solution:**

$$\begin{aligned} L(\theta) &= \prod \frac{\theta}{x_i^{\theta+1}} \\ l(\theta) &= n\ln(\theta) - (\theta+1)\sum \ln(x_i) \\ l'(\theta) &= \frac{n}{\theta} - \sum \ln(x_i) \\ \hat{\theta} &= \frac{n}{\sum \ln(x_i)} = \frac{6}{11.3861} = 0.5270 \end{aligned}$$

(b) [**3 marks**] Compute an approximate 95% confidence interval for $\theta$. Comment on the appropriateness of your confidence interval.

**Solution:**

$$\begin{aligned} l''(\theta) &= -\frac{n}{\theta^2} \\ \hat{\theta} &\sim N(\theta, \frac{\theta^2}{n}) \\ CI &= \hat{\theta} \pm 2\sqrt{\frac{\hat{\theta}^2}{n}} \\ &= 0.5270 \pm 0.4303 = (0.0967, 0.9573) \end{aligned}$$

This confidence interval is not that appropriate because the sample size is small.

(c) [**4 marks**] Estimate the probability that an insect will survive for more than 2 days. Provide a standard error for your estimate.

**Solution:**

$$
\begin{aligned}
\hat{S}(x) =& \frac{1}{x^{\hat{\theta}}} \\
g(\hat{\theta}) =& \hat{S}(2) = 2^{-\hat{\theta}} = 0.6940 \\
g'(\hat{\theta}) =& -2^{-\hat{\theta}} \ln 2 \\
Var(g(\hat{\theta})) \approx& Var(\hat{\theta})(g'(\hat{\theta}))^2 \\
=& \frac{0.5270^2}{6}(2^{-0.5270} \ln 2)^2 \\
=& 0.010711 \\
S.E =& 0.1035
\end{aligned}
$$

# END OF EXAMINATION