A.3 Life expectancy, Lexis diagrams and census approximation

1. Show how to represent on a Lexis diagram (i) the number of individuals in a population who are aged 20 at the start of 2014; (ii) the number of individuals in the population who turn 21 during 2014.

2. Consider the situation of estimating mortality rates by observing a population over the course of a time interval $[K, K+N]$.

Let $d_x^{(1)}$ be the number of deaths at curtate age $x$, and $E_x^{c(1)}$ the corresponding central exposure to risk, given by

$$E_x^{c(1)} = \int_K^{K+N} P_x(t) dt,$$

where $P_x(t)$ is the number of individuals under observation at time $t$ with curtate age $x$. If we observe only census data at times $K, K + 1, \ldots, K+N$, we can use linear interpolation to approximate

$$E_x^{c(1)} \approx \sum_{k=K+1}^{K+N} \frac{1}{2} \left( P_{x,k-1}^{(1)} + P_{x,k}^{(1)} \right).$$

Then $d_x^{(1)}/E_x^{c(1)}$ gives an estimator of $\mu_{x+1/2}$, under the assumption that the mortality rate $\mu_t$ at age $t$ is constant over $t \in [x, x+1]$.

More generally, $d_x^{(1)}/E_x^{c(1)}$ will approximate $\int_{t=x}^{x+1} \mu_t dt$ as long as the number observed does not change too significantly over the relevant time (say, if mortality is low and there are not significant numbers of people entering or leaving the observed population for other reasons).

(a) State the Principle of Correspondence. Adapt the properties above to the following different definitions of $d_x$ (giving corresponding definitions and suitable approximations for $E_x^{c}$, and explaining what may be estimated by $d_x/E_x^{c}$). Where appropriate, explain what further assumptions are needed:

- $d_x^{(2)} = \#$ deaths with $x$ nearest birthday to death;
- $d_x^{(3)} = \#$ deaths in calendar year of the $x$th birthday;
- $d_x^{(4)} = \#$ deaths with curtate age $x$ at time of last annual policy renewal;

(b) For cases of $d_x^{(1)}, d_x^{(2)}$ and $d_x^{(3)}$, indicate the areas relevant to the calculation of deaths and exposure on a Lexis diagram.

3. In the setting of Question 2., assume that both birthdays and policy anniversaries are uniformly spread over the year. Consider

$$d_x^{(5)} = \# \text{ deaths with last annual policy renewal in the calendar year of } x\text{th birthday.}$$

(a) Show that the possible age range of deaths counted under $d_x^{(5)}$ is $(x-1, x+2)$.

(b) For each $y \in (x-1, x+2)$, calculate the probability that a death aged $y$ is counted by $d_x^{(5)}$.

(c) Using the Principal of Correspondence, specify the denominator $E_x^{c(5)}$ for which $d_x^{(5)}/E_x^{c(5)}$ should be close to a weighted integral $\int_{-1}^{2} g(u)\mu_{x+u} du$ over the force of mortality between $x - 1$ and $x + 2$. Specify the weight function $g$. 

4. (a) Suppose you are given estimates for a population of remaining life expectancy $\hat{e}_x$ and $\hat{e}_{x+t}$, corresponding to ages $x$ and $x+t$ (years). You wish to compute the mortality probability $tq_x$. Under the assumption that mortality rates are constant over this interval, explain how to derive the approximation

$$tq_x \approx \frac{t+\hat{e}_{x+t} - \hat{e}_x}{t/2 + \hat{e}_{x+t}}.$$  

Under what conditions will this approximation be reasonable?

(b) The following is an estimated table of $\hat{e}_x$ (in years) in ancient Rome, as computed by Tim Parkin *Demography and Roman Society* (Johns Hopkins University Press, 1992).

<table>
<thead>
<tr>
<th>$x$</th>
<th>0</th>
<th>1</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
<th>35</th>
<th>40</th>
<th>45</th>
<th>50</th>
<th>55</th>
<th>60</th>
<th>65</th>
<th>70</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{e}_x$</td>
<td>25</td>
<td>33</td>
<td>43</td>
<td>41</td>
<td>37</td>
<td>34</td>
<td>32</td>
<td>29</td>
<td>26</td>
<td>23</td>
<td>20</td>
<td>17</td>
<td>14</td>
<td>10</td>
<td>8</td>
<td>6</td>
</tr>
</tbody>
</table>

On the basis of these figures, and assuming the mortality rates to be constant over the relevant age intervals, use equation (*) to approximate the annual mortality probabilities $1q_x$ over the age intervals 0–1, 1–5, 5–10.

(c) Suppose we know that 15% of Roman infants died of dysentry in their first year. Under the competing risks assumption, estimate the change in life expectancy at birth that would have resulted if this disease had been eliminated among infants. Note the assumptions you make in carrying out the calculation.

5. A large investigation has been carried out into mortality among people of working age, recording Observed deaths $d_x$ in 5-year intervals and associated Exposed to risk $E_x$ (5-year sums of annual initial exposed to risk). They are to be compared with a well-known standard table of one-year death probabilities $q^s_x$, constant within the 5-year intervals.

<table>
<thead>
<tr>
<th>$x$</th>
<th>$E_x$</th>
<th>$d_x$</th>
<th>$q^s_x \times 10^5$</th>
<th>$x$</th>
<th>$E_x$</th>
<th>$d_x$</th>
<th>$q^s_x \times 10^5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20–24</td>
<td>35000</td>
<td>35</td>
<td>97</td>
<td>45–49</td>
<td>28000</td>
<td>138</td>
<td>460</td>
</tr>
<tr>
<td>25–29</td>
<td>33000</td>
<td>30</td>
<td>88</td>
<td>50–54</td>
<td>25000</td>
<td>229</td>
<td>850</td>
</tr>
<tr>
<td>30–34</td>
<td>30000</td>
<td>31</td>
<td>117</td>
<td>55–59</td>
<td>23000</td>
<td>360</td>
<td>1500</td>
</tr>
<tr>
<td>35–39</td>
<td>30000</td>
<td>45</td>
<td>173</td>
<td>60–64</td>
<td>20000</td>
<td>522</td>
<td>2500</td>
</tr>
<tr>
<td>40–44</td>
<td>31000</td>
<td>84</td>
<td>260</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Perform the following three tests, finding the $p$-values and the test statistic (where appropriate): (a) $\chi^2$-test (b) sign test (c) cumulative-deviations test, commenting on the outcomes.