Faculty of Actuaries **Institute of Actuaries**

REPORT OF THE BOARD OF EXAMINERS

September 2003

Subject 104 — Survival Models

EXAMINERS' REPORT

Introduction

The attached subject report has been written by the Principal Examiner with the aim of helping candidates. The questions and comments are based around Core Reading as the interpretation of the syllabus to which the examiners are working. They have however given credit for any alternative approach or interpretation which they consider to be reasonable.

J Curtis
Chairman of the Board of Examiners

11 November 2003

EXAMINATIONS

September 2003

Subject 104 — Survival Models

EXAMINERS' REPORT

EXAMINERS' COMMENTS

As in previous years, the Examiners aimed to set questions covering all the aspects of Survival Modelling: Life contingencies including its stochastic treatment, Graduation including its statistical aspects and the determination of exposures. The Examiners aim to strike a balance between questions requiring numerical solutions and those requiring verbal and algebraic answers, as well as between those with and without a statistical theme.

Comments on solutions presented to individual questions for this September 2003 paper are given below:

Question 1: *Many candidates failed to give sufficient distinct points to score full credit.*

Question 2: *This was poorly answered. Candidates were asked to "show" and therefore all steps needed to be clearly shown for full credit. Some candidates attempted to derive the "variance adjustment factor" given in the Gold book, not the expression given in the question.*

Question 3: *Part (i) was well answered.*
In (ii), few candidates adequately explained how they reached the correct answer, even though a derivation was required, and therefore failed to score full marks. Some candidates used 1 January 2002 rather than 1 July 2001.

Question 4: *Part (i) was well answered, with most candidates attempting this and scoring highly. A reasonably common error was incorrect inequalities on the ranges for t.*
Part (ii) was poorly answered. Most candidates plotted a step function, with very few correctly identifying what the hazard was. Many plotted S(t) or F(t). Some attempted to use the Nelson-Aalen estimate, using $h(t) = -\ln S(t)$.

Question 5: *Most candidates correctly identified the rate interval in part (i).*
In part (ii), frequent errors were looking at only one year (so no summation) and stating the wrong limits on the summation. Few candidates stated the correct assumption.
Part (iii) was generally not well answered. Few candidates correctly identified the age to which $\hat{\mu}_x$ *applied, with a large number simply stating*

$\hat{\mu}_{60} = \dfrac{d_{60}}{E_{60}^c}$ *without any further consideration.*

Question 6: *Overall this was well answered, although many candidates struggled to show clearly enough part (ii)(a). Most candidates scored full credit for (iii).*

Question 7: *Parts (i) and (ii) were well answered, although many candidates did not use the data in the question.*
The quality of the answers to part (iii) was more varied. Many candidates assumed $E_{71}^c = \frac{1}{2}\left(P_{71,0} + P_{72,1}\right)$. *Some candidates confused* E_x^c *and* E_x.

Question 8: *Part (i)(a) was correctly answered by the majority of the candidates. Many candidates struggled with part (i)(b), failing to realise the effect this type of mortality loading has on the calculations, even though this has been examined several times in recent years. Many candidates then failed to produce an answer for part (ii). Some credit was given to those candidates who correctly identified how to calculate the expected present value of the profit, even if the numerical answer was not correct.*

Question 9: *This was very poorly answered. Many candidates did not make a serious attempt at this question. Few candidates gave the correct relationship in part (i). In part (iii), many solutions lacked the level of detail required to show Thiele's equation, as requested. Questions on this subject in previous years were also poorly answered.*

Question 10: *In part (i), a surprising number of candidates did not correctly state both dates. The most common error was to give the final date as 2014 (possibly by adding 5 to the correct first date of 2009), which would mean 6 payments. Part (ii) and (iii) were intended to lead candidates through the calculation of the premium required in part (iv), but few candidates realised this and these parts were very poorly answered. Many candidates failed to realise that* K_x *and X were random variables and included probabilities of survival or death in part (ii), then failing to answer part (iii) sensibly. Many plots were shown as continuous rather than discrete. In part (iv) a surprising number of candidates calculated the correct premium, even though they had not answered earlier parts correctly. Credit was given for this, although it did highlight the lack of understanding about parts (ii) and (iii).*

Question 11: *Many candidates made reasonable attempts at this question and scored well on part (ii). Candidates lost marks by giving insufficient detail and explanation of the tests they were carrying out and by not stating their conclusions clearly.*

1 It is normally assumed that the true mortality rates in the population under investigation are smooth functions of age.

Crude mortality rates are typically estimated separately at each integer age, and therefore may not progress smoothly.

Graduation allows information from adjacent ages to be used to improve the estimate at each age, thus reducing sampling errors.

It is desirable that financial quantities progress smoothly with age, as irregularities are hard to justify in practice. If the underlying mortality rates are smooth, then financial quantities calculated using them will also be smooth.

2 Let D_i denote the number of deaths among the $\pi_i N$ lives with i policies, and let C_i denote the number of claims among the same group. Then $C_i = iD_i$ and

$$D = \sum_i D_i$$

$$C = \sum_i C_i = \sum_i iD_i.$$

Now $D_i \sim B(\pi_i N, q_x)$ since lives are genuinely independent and

$$\begin{aligned}\text{Var}(C) &= \text{Var}\left(\sum_i iD_i\right) \\ &= \sum_i i^2 \text{Var}(D_i) \\ &= \sum_i i^2 \pi_i N q_x (1 - q_x)\end{aligned}$$

and

$$\begin{aligned}\text{Var}(D) &= \text{Var}\left(\sum_i D_i\right) \\ &= \sum_i \pi_i N q_x (1 - q_x)\end{aligned}$$

so

$$\frac{\mathrm{Var}(C)}{\mathrm{Var}(D)} = \frac{\Sigma_i i^2 \pi_i N q_x (1-q_x)}{\Sigma_i \pi_i N q_x (1-q_x)}$$

$$= \frac{\Sigma_i i^2 \pi_i}{\Sigma_i \pi_i}.$$

3 (i) For each life in the investigation we require:

date of birth (or date of xth birthday);
date of entry into observation;
date of exit from observation.

(ii) The central exposed to risk, E_x^c is given by the formula

$$E_x^c = \int_0^2 P_{x,t} dt ,$$

where $P_{x,t}$ is the number of lives under observation aged x last birthday at time t, measured as the duration since the start of the investigation.

To estimate this, we use the trapezium rule (assuming $P_{x,t}$ is linear between census dates).

Let the population aged x last birthday on 1 January 2001 be $P_{x,0}$; and the corresponding populations on 1 July 2001 and 1 January 2003 be $P_{x,0.5}$ and $P_{x,2}$.

The exposed to risk for the period 1 January–1 July 2001 is 0.5 $(P_{x,0} + P_{x,0.5}) \times$ length of period in years

$= 0.5\ (P_{x,0} + P_{x,0.5}) \times 0.5.$

Similarly, the exposed to risk for the period 1 July 2001–1 January 2003 is
0.5 $(P_{x,0.5} + P_{x,2}) \times$ length of period in years
$= 0.5\ (P_{x,0.5} + P_{x,2}) \times 1.5.$

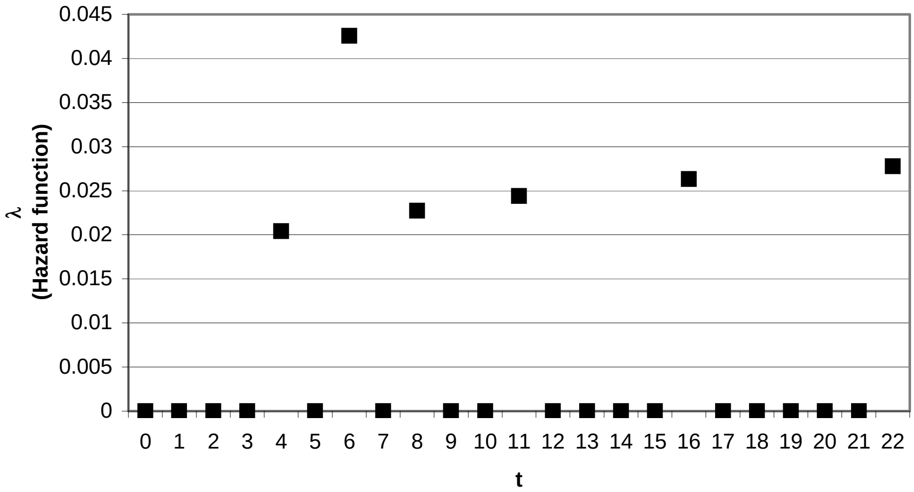
Therefore the total exposed to risk for the entire period is
$[0.5\,(P_{x,0} + P_{x,0.5}) \times 0.5] + [0.5\,(P_{x,0.5} + P_{x,2}) \times 1.5]$
$= 0.25\,P_{x,0} + P_{x,0.5} + 0.75\,P_{x,2}$.

4 (i) $S(t) = \prod_{t_j < t} (1 - \hat{\lambda}_j)$

t_j	c_j	d_j	n_j	$\hat{\lambda}_j$
0	1	0	50	0
4	1	1	49	1/49
6	1	2	47	2/47
8	2	1	44	1/44
11	2	1	41	1/41
16	1	1	38	1/38
22	0	1	36	1/36

t	$S(t)$
$0 \le t < 4$	1
$4 \le t < 6$	0.9796
$6 \le t < 8$	0.9379
$8 \le t < 11$	0.9166
$11 \le t < 16$	0.8942
$16 \le t < 22$	0.8707
$t \ge 22$	0.8465

(ii)



5 (i) (a) The age definition of the death data is:

x = age last birthday at date of issue plus number of complete years since issue
= age last birthday at previous policy anniversary

Age changes from x to $x+1$ on a policy anniversary.
This is therefore a policy year rate interval.

(b) $$E_x^c = \int_0^N P_{x,t}\,dt = \sum_{t=0}^{N-1} \frac{1}{2}\left(P_{x,t} + P_{x,t+1}\right)$$

Assuming policy anniversaries are uniformly spread over the calendar year.

(ii) $\hat{\mu}_x = \dfrac{d_x}{E_x^c}$ estimates μ at the mean age halfway through the interval.

Assuming birthdays are uniformly distributed over policy years,
the mean age at the start of the interval is $x + ½$.
The mean age halfway though is then $x + 1$.

So, $\hat{\mu}_x$ estimates μ_{x+1}

So, $\hat{\mu}_{59} = \dfrac{d_{59}}{E_{59}^c} = \dfrac{75}{\frac{1}{2}(6276 + 6824)} = 0.01145$ estimates μ_{60}.

(iii) Premiums are based on age last birthday. There could be a tendency for policyholders to take out policies just before their birthday to benefit from lower premiums. The assumption that birthdays are uniformly distributed would then be invalid.

Other plausible reasons why assumptions made may not be valid were given credit.

6 (i) $f_x(t) = {}_t p_x \cdot \mu_{x+t}$

(ii) (a) Note that ${}_s q_x = \int_0^s f_x(r)\,dr$

$$\Rightarrow \frac{\partial}{\partial s}\,{}_s p_x = -\frac{\partial}{\partial s}\,{}_s q_x = -f_x(s) = -\,{}_s p_x \mu_{x+s}$$

and $\dfrac{\partial}{\partial s}\log {}_s p_x = \dfrac{\frac{\partial}{\partial s}\,{}_s p_x}{{}_s p_x}$

So, $\dfrac{\partial}{\partial s}\log {}_s p_x = -\mu_{x+s}$

(b) Hence, $\int_0^t \frac{\partial}{\partial s}\log {}_s p_x\,ds = -\int_0^t \mu_{x+s}\,ds$

$$\Rightarrow \left[\log {}_s p_x\right]_0^t = \log {}_t p_x = -\int_0^t \mu_{x+s}\,ds \quad (\text{since } {}_0 p_x = 1)$$

taking exponentials of both sides gives,

$${}_t p_x = \exp\left\{-\int_0^t \mu_{x+s}\,ds\right\} \text{ as required}$$

(iii) Required probability is ${}_{15|3}q_{65} = {}_{15}p_{65} \cdot {}_3 q_{80} = {}_{15}p_{65} \cdot \left(1 - {}_3 p_{80}\right)$

$${}_{15}p_{65} = \exp\left\{-\int_0^5 0.01\,ds - \int_5^{15} 0.015\,ds\right\}$$

$$= \exp\left\{-(5 \times 0.01) - (10 \times 0.015)\right\} = e^{-0.2}$$

$${}_3 p_{80} = \exp\left\{-\int_0^3 0.025\,ds\right\} = \exp\{-3 \times 0.025\} = e^{-0.075}$$

Required probability = $e^{-0.2}\,(1 - e^{-0.075}) = 0.059$

7 (i) ${}_t p_x^{RD}$ is the probability that a life aged x and in the retired state at time 0 will be in the dead state at time t.

(ii) (a) Omitting the x subscripts for clarity, the likelihood function is given by

$$L = \text{Const} \times \prod_{i=1}^{N} e^{-(v+\rho)(b_i-a_i)} \times v^M \rho^K.$$

This solution assumes (as was intended by the question) that employees ceased to be observed when they retired. Credit was also given if candidates assumed observation continued into retirement, so that deaths after retirement were taken into account.

(b) The log-likelihood is then given by

$$\log L = \text{const} + \sum_{i=1}^{N} -(v+\rho)(b_i - a_i) + M \log v + K \log \rho.$$

Differentiating to find the maximum gives

$$\frac{d}{d\rho} \log L = -\sum_{i=1}^{N} (b_i - a_i) + \frac{K}{\rho}.$$

To find the maximum likelihood estimate, we set this equal to zero so

$$\frac{K}{\hat{\rho}} = \sum_{i=1}^{N} (b_i - a_i)$$

so

$$\hat{\rho} = \frac{K}{\sum_{i=1}^{N} (b_i - a_i)}.$$

We check this is a maximum by finding the second derivative:

$$\frac{d^2}{d\rho^2} \log L = -\frac{K}{\rho^2} < 0$$

so indeed we have a maximum.

(iii) $E_{71}^c \simeq 0.5 \times (P_{71}(2001) + P_{71}(2002))$

$\approx 0.5 \times (7592 + 8062) = 7827$

so $\hat{\mu}_{71} = \dfrac{d_{71}}{E_{71}^c} = \dfrac{76}{7827} = 0.009710$

and $q_{71} \approx 1 - e^{-\hat{\mu}_{71}}$
$= 0.009663.$

Similarly

$$E_{72}^c \approx 0.5 \times (6811 + 7493) = 7152$$

and

$$\hat{\mu}_{72} = \frac{107}{7152} = 0.01496.$$

and hence

$$q_{72} \approx 1 - e^{-\hat{\mu}_{72}} = 0.01485.$$

Assumptions

- μ is constant over year of age
- population varies linearly between census dates.

ALTERNATIVELY

(iii) $E_{71}^C \simeq 0.5 \times (7592 + 8062) = 7827$ as above

$$E_{71} = E_{71}^c + \frac{d_{71}}{2}$$

$$= 7827 + \frac{76}{2} = 7865$$

so $q_{71} = \dfrac{d_{71}}{E_{71}}$

$$= \frac{76}{7865} = 0.009663$$

Similarly,

$$E_{72}^{c} \simeq 0.5 \times (6811 + 7493) = 7152$$

$$E_{72} = 7152 + \frac{107}{2} \qquad = 7205.5$$

$$q_{72} = \frac{107}{7205.5} \qquad = 0.01485$$

Assumptions

- birthdays are distributed uniformly over calendar years
- deaths are distributed uniformly over calendar years

8 (i)

(a) Equation of value: $P = X \cdot {}_{15|}\bar{a}_{50} = X \cdot \frac{D_{65}}{D_{50}} \bar{a}_{65} = X \cdot \frac{D_{65}}{D_{50}} \left(\ddot{a}_{65} - 0.5\right)$

$$\Rightarrow 20000 = X \cdot \frac{689.23}{1366.61} \cdot \left(12.276 - 0.5\right) = 5.939055X$$

$\Rightarrow X =$ £3,367.54 per annum

(b) Using the special mortality rates,

$${}_t p_x^S = \exp\left\{-\int_0^t \left(\mu_{x+s} + 0.019048\right) ds\right\}$$

$$= \exp\left\{-0.019048t - \int_0^t \mu_{x+s} ds\right\} = e^{-0.019048t} \cdot {}_t p_x$$

So, $\frac{D_{x+t}^S}{D_x^S} = v^t \cdot {}_t p_x{}^s = v^t \cdot e^{-0.019048t} \cdot {}_t p_x = \frac{D_{x+t}}{D_x}$ calculated at rate of interest j such that $(1 + j)^{-1} = v.e^{-0.019048}$ => $j = 6\%$.

Or $\frac{D_{x+t}^S}{D_x^S} = e^{-0.019048t} \cdot \frac{D_{x+t}}{D_x}$ calculated at 4%

and

$$\bar{a}_x^S = \int_0^\infty v^t \cdot {}_t p_x^S dt = \int_0^\infty v^t \left(e^{-0.19048t} \cdot {}_t p_x\right) dt = \int_0^\infty \left(v \cdot e^{-0.19048}\right)^t \cdot {}_t p_x dt$$

$= \bar{a}_x$ calculated at rate of interest 6%

$$\text{So, } 20000 = X \cdot \frac{1.06^{-65} \times 8821.2612}{1.06^{-50} \times 9712.0728} \cdot (10.569 - 0.5) = 3.81608X$$

$\Rightarrow X = £5{,}240.98$ per annum

$$\text{Or } 20000 = X \cdot e^{-0.28572} \cdot \frac{689.23}{1366.61} \cdot (10.569 - 0.5) = 3.81610X$$

$\Rightarrow X = £5{,}240.96$ per annum

(ii) EPV Profit= EPV of income – EPV of outgo

$$= 20000 - 3367.54 \times \frac{D_{65}^S}{D_{50}^S} \times \bar{a}_{65}$$

$$= 20000 - 3367.54 \times 3.81610$$

$$= £7{,}149.20$$

9 (i) $e^{\delta}(V_t + P) = p_{60+t}(V_{t+1} + 1)$ $\quad t = 0$

$e^{\delta}V_t = p_{60+t}(V_{t+1} + 1)$ $\quad t > 0$

Explanation: LHS = reserve at time *t* plus one year's interest. This must equal the reserve needed at time *t* + 1, plus the benefit payment then due, allowing for the probability that the policyholder survives from time *t* to time *t* + 1, which is the RHS.

(ii) $e^{\delta h}V_t = {}_hp_{60+t}(V_{t+h} + h)$.

(iii) Using $e^{\delta h} = 1 + \delta h + o(h)$

and ${}_hp_{60+t} = 1 - {}_hq_{60+t} = 1 - h\mu_{60+t} + o(h)$

Rewriting the above gives

$$[1 + \delta h + o(h)]\, V_t = (1 - h\mu_{60+t} + o(h))\,(V_{t+h} + h)$$

so $$V_{t+h} - V_t = \delta h V_t - h + h\mu_{60+t}\, V_{t+h} + o(h)$$

so $$\frac{V_{t+h} - V_t}{h} = \delta V_t + \mu_{60+t} V_{t+h} - 1 + o(h)/h$$

Taking the limit as $h \to 0$ gives

$$\frac{\partial V_t}{\partial t} = (\delta + \mu_{60+t})\, V_t - 1$$

as required. The boundary condition is

$$V_{20} = 0.$$

(iv) We have

$$\frac{\partial V_t}{\partial t} = (\delta + \mu)\, V_t - 1$$

so $$\frac{\partial V_t}{\partial t} - (\delta + \mu)\, V_t = -1$$

so $$e^{-(\delta+\mu)t}\,\frac{\partial V_t}{\partial t} - (\delta + \mu)e^{-(\delta+\mu)t}V_t = -\,e^{-(\delta+\mu)t}.$$

This means that

$$\frac{\partial}{\partial t}[e^{-(\delta+\mu)t}V_t] = -\,e^{-(\delta+\mu)t}$$

so $$e^{-(\delta+\mu)t}V_t = \int -e^{-(\delta+\mu)t}\,dt \qquad (1)$$

$$= \frac{1}{\delta+\mu}e^{-(\delta+\mu)t} + C \qquad (2)$$

Hence

$$V_t = \frac{1}{\delta+\mu} + Ce^{(\delta+\mu)t}.$$

Using the boundary condition $V_{20} = 0$ we have

$$0 = \frac{1}{\delta+\mu} + Ce^{(\delta+\mu)20}$$

so $\quad C = -\dfrac{e^{-(\delta+\mu)20}}{\delta+\mu}.$

So the solution is

$$V_t = \frac{1}{\delta+\mu}[1 - e^{-(\delta+\mu)20}e^{(\delta+\mu)t}] \quad (3)$$

$$= \frac{1}{\delta+\mu}[1 - e^{-(\delta+\mu)(20-t)}] \quad (4)$$

$$= \bar{a}_{\overline{20-t}|} \quad (5)$$

where the annuity is calculated using an effective force of interest of $\delta + \mu$.

10 (i) First payment is on 31 August 2009 (1 September 2009 acceptable).

Last payment is on 31 August 2013 (1 September 2013 acceptable).

(ii) For $K_x \leqslant 5$, present value (PV) = 0.

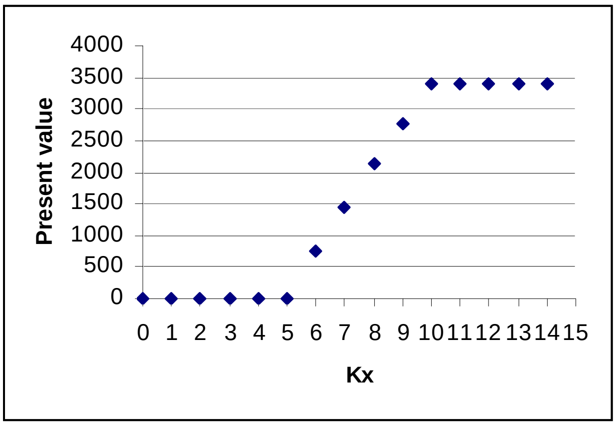
For $6 \leqslant K_x \leqslant 10$, PV = $1000\sum_{k=6}^{K_x} v^k$.

For $K_x > 10$, PV = $1000\sum_{k=6}^{10} v^k$.

Hence, using $v = \dfrac{1}{1.05}$ we have

K_x	*present value*
$0 \leqslant 5$	0
6	746.22
7	1,456.90
8	2,133.74
9	2,778.35
$\geqslant 10$	3,392.26

The plot is shown below.



Credit was also given to candidates who used the approximation given in the Gold book.

(iii) The calculations are shown below, using AM92 ultimate mortality.

$$\text{P}[\mathbf{X} = 0] = {}_6q_{60} = 1 - \frac{l_{66}}{l_{60}} = 1 - \frac{8695.6199}{9287.2164} = 0.06370$$

$$\text{P}[\mathbf{X} = 746.22] = {}_6p_{60}.q_{66} = \frac{l_{66} - l_{67}}{l_{60}} = \frac{d_{66}}{l_{60}} = \frac{138.6082}{9287.2164} = 0.01492$$

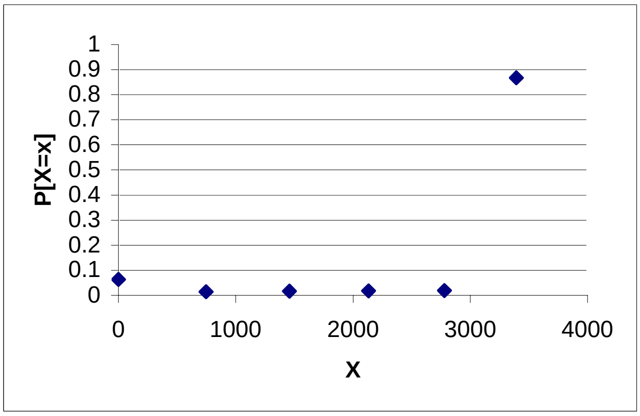
$$\text{P}[\mathbf{X} = 1{,}456.90] = {}_7p_{60}.q_{67} = \frac{l_{67} - l_{68}}{l_{60}} = \frac{d_{67}}{l_{60}} = \frac{152.5202}{9287.2164} = 0.01642$$

$$\text{P}[\mathbf{X} = 2{,}133.73] = {}_8p_{60}.q_{68} = \frac{l_{68} - l_{69}}{l_{60}} = \frac{d_{68}}{l_{60}} = \frac{167.3586}{9287.2164} = 0.01802$$

$$\text{P}[\mathbf{X} = 2{,}778.35] = {}_9p_{60}.q_{69} = \frac{l_{69} - l_{70}}{l_{60}} = \frac{d_{69}}{l_{60}} = \frac{183.0785}{9287.2164} = 0.01971$$

$$\text{P}[\mathbf{X} = 3{,}392.26] = {}_{10}p_{60} = \frac{l_{70}}{l_{60}} = \frac{8054.0544}{9287.2164} = 0.86722$$

The plot is shown below.



(iv) Single premium = expected present value of benefits
Using the data from part (iii) we have

EPV benefits = $\sum x.(\text{P}[\text{X} = x])$

The calculations are shown in the table below

x	Pr[$\mathbf{X} = x$]	x. Pr[$\mathbf{X} = x$]
0	0.06370	0
746.22	0.01492	11.1336
1,456.90	0.01642	23.9223
2,133.74	0.01802	38.4500
2,778.35	0.01971	54.7613
3,392.26	0.86722	2941.8357
	sum	3,070.1029

Therefore the premium required is £3,070.10.

11 (i) On average, the underlying mortality of the lives in the investigation could be systematically heavier or lighter than that represented by the standard table.

Even if, overall, the mortality in the investigation is not significantly different from that in the standard table, there could be individual age-groups where large differences exist.

Even if, overall, the mortality in the investigation is not significantly different from that in the standard table, there could be significant sections of the age range (i.e. runs of consecutive age groups) over which it is heavier or lighter.

Reasons as to <u>why</u> these points might be the case (for example discussions about homogeneity) were given credit if valid in the context of the question.

(ii) The null hypothesis for all the tests is that the underlying mortality of the lives in the investigation is that of the standard table.

CHI SQUARED TEST

The Chi-squared test is based on z_x^2, the calculation of which is shown below

Age group	*Actual deaths*	*Expected deaths*	z_x	z_x^2
20–24	35	34	0.17150	0.02941
25–29	30	29	0.18569	0.03448
30–34	31	35	−0.67612	0.45714
35–39	45	52	−0.97072	0.94231
40–44	84	80	0.44721	0.20000
45–49	138	130	0.70165	0.49231
50–54	229	213	1.09630	1.20188
55–59	360	348	0.64327	0.41379
60–64	522	505	0.75649	0.57228

Using the data in the table above, $\sum_x z_x^2 = 4.34360$.

This is a test of overall adherence of the data to the standard table.

The test statistic is $\sum_x z_x^2 \sim \chi_m^2$,

where m is the number of age groups ($m = 9$ in our case), because we are comparing an experience with a standard table.

This is less than the critical value of the χ_9^2 distribution at the 5% level.

We accept the null hypothesis.

STANDARDISED DEVIATIONS TEST

This tests for the possibility that there are a small number of age groups with large differences between the mortality rates in the investigation and the standard table.

The z_xs comprise m independent samples from a Normal (0,1) distribution. We can compare the expected and actual number of deviations in the following ranges:

Range	(-3,-2)	(2,-1)	(-1,0)	(0,1)	(1,2)	(2,3)
Expected	0.18	1.26	3.06	3.06	1.26	0.18
Actual	0	0	2	6	1	0

Therefore under the null hypothesis we should expect fewer than 1 in 20 to be > 2 in absolute magnitude. In this case none of the z_xs exceeds 2 in absolute value, so we accept the null hypothesis.

Also, under the null hypothesis about half the deviations will lie between -2/3 and +2/3. In this case 4 out of 9 do, which is consistent with the null hypothesis.

SIGNS TEST

This tests for the possibility of the mortality rates in the investigation being systematically lower or higher than those in the standard table.

Let P be the number of z_xs that are positive.

Then under the null hypothesis, P ~ Binomial(9, 0.5).

We have 7 positive signs. The probability of getting 7 or more positive signs if the null hypothesis is true is (also available from tables in Gold book):

$$\begin{aligned}
&\binom{9}{7}\frac{1}{2^9}+\binom{9}{8}\frac{1}{2^9}+\binom{9}{9}\frac{1}{2^9}\\
&=\frac{9!}{7!2!}\frac{1}{2^9}+\frac{9!}{8!1!}\frac{1}{2^9}+\frac{1}{2^9}\\
&=(36+9+1)0.001953125\\
&=0.08984375.
\end{aligned}$$

Or in this case, it is sufficient to look at the probability of getting just 7 signs (= 0.0703).

This is greater than 0.025 (2-tailed test)

We accept the null hypothesis.

CUMULATIVE DEVIATIONS TEST

When using the whole age range, this tests for the possibility of the mortality rates in the investigation being systematically lower or higher than those in the standard table.

Under the null hypothesis,

$$\frac{\sum_{x=1}^{m}(d_x - E_x q_x^s)}{\sqrt{\sum_{x=1}^{m} E_x q_x^s}} \sim \text{Normal}(0,1)$$

Using the data in the table, we have

$$\frac{\sum_{x=1}^{m}(d_x - E_x q_x^s)}{\sqrt{\sum_{x=1}^{m} E_x q_x^s}} = \frac{48}{\sqrt{1{,}426}} = 1.271.$$

Since both positive and negative cumulative deviations are of interest we use a two-tailed test.

Since $|1.271|<1.96$ we accept the null hypothesis.

GROUPING OF SIGNS TEST

This tests for runs of deviations of the same sign, that is for subsections of the age range for which the mortality rates of lives in the investigation are systematically lower or higher than the rates in the standard table.
Let G be the number of groups of positive z_xs.

Let n_1 be the number of positive z_xs and n_2 be the number of negative z_xs.

In our case $G = 2$, $n_1 = 7$ and $n_2 = 2$.

Then the probability of getting 2 or fewer groups of positive signs is

$$\frac{\binom{6}{1}\binom{3}{2}}{\binom{9}{7}}+\frac{\binom{6}{0}\binom{3}{1}}{\binom{9}{7}}=\frac{\frac{6!}{1!5!}\frac{3!}{1!2!}+\frac{3!}{2!1!}}{\frac{9!}{7!2!}}=\frac{21}{36}=0.58333$$

Using a one-tailed test, since only small values of G are of interest, we find that $0.58333 > 0.05$.

We accept the null hypothesis.

SERIAL CORRELATIONS TEST

This tests for runs of deviations of the same sign, that is for subsections of the age range for which the mortality rates of lives in the investigation are systematically lower or higher than the rates in the standard table.

The correlation coefficient at lag 1 is

$$r_1 = \frac{\sum_{x-1}^{m-1}(z_x - \overline{z^*})(z_{x+1} - \overline{z^{**}})}{\sqrt{\sum_{x=1}^{m-1}(z_x - \overline{z^*})^2 \sum_{x=1}^{m-1}(z_{x+1} - \overline{z^{**}})^2}}. \qquad +½$$

where $\overline{z^*}$ is the mean of the standard deviations from ages 1 to $m - 1$ and $\overline{z^{**}}$ is the mean of the standard deviations from ages 2 to m.

The calculations are shown in the table below.

Age group	z_x	$z_x - \overline{z^*}$	$z_x - \overline{z^{**}}$
20–24	0.17150	–0.028348	–0.10147
25–29	0.18569	–0.014158	–0.08728
30–34	–0.67612	–0.875968	–0.94909
35–39	–0.97072	–1.170568	–1.24369
40–44	0.44721	0.247363	0.17424
45–49	0.70165	0.501803	0.42868
50–54	1.09630	0.896453	0.82333
55–59	0.64327	0.443423	0.3703
60–64	0.75649	0.55664	0.48352

$\overline{z^*} = 0.19985$

$\overline{z^{**}} = 0.27297$

Age group	$(z_x - \overline{z^*})^2$	$(z_x - \overline{z^{**}})^2$	$(z_x - \overline{z^*})(z_{x+1} - \overline{z^{**}})$
20–24	0.000804	0.010296	0.00247
25–29	0.000200	0.007618	0.013437
30–34	0.767319	0.90077	1.089433
35–39	1.370228	1.546765	–0.203960
40–44	0.061188	0.030360	0.106040
45–49	0.251806	0.183767	0.413149
50–54	0.803627	0.67787	0.331957
55–59	0.196624	0.137122	0.214404
60–64	0.309848	0.23379	

Hence

$$r_1 = \frac{\sum_{x-1}^{m-1} (z_x - \overline{z^*})(z_{x+1} - \overline{z^{**}})}{\sqrt{\sum_{x=1}^{m-1} (z_x - \overline{z^*})^2 \sum_{x=1}^{m-1} (z_{x+1} - \overline{z^{**}})^2}} = \frac{1.96693}{\sqrt{(3.451796)(3.718062)}} = 0.549045.$$

Now $r_1 \rfloor m \sim$ Normal(0,1).

Since $m = 9$, we have $r_1 \rfloor m = 3 \times 0.549045 = 1.64714$.

Using a one-tailed test (since we are only interested in positive serial correlations), the probability of getting a value as high as 1.64714 is 0.05.

Therefore we have just sufficient evidence to reject the null hypothesis.

This test uses the exact formula, as given in the Core Reading. Credit was also given to candidates who used the approximation given in the Gold book.

(iii) There is little evidence here to suggest that the mortality of the lives in the investigation is significantly different from that represented by the standard table.

The experience passes all the tests except the serial correlations test (and only fails that marginally).

However, there is a suggestion that at ages over 40, mortality is consistently heavier than that in the standard table.

Moreover it is at ages over 40 that most deaths occur.

Therefore life offices using the standard table to represent the mortality experience of lives such as those in the investigation might find their profitability impaired, as they would tend to charge premiums which are too low.