EXAMINATIONS

September 2000

Subject 104 — Survival Models

EXAMINERS’ REPORT

General Comment

In many cases a lack of statistical and mathematical knowledge prevented candidates from completing many questions satisfactorily.

For example the failure to integrate simple exponenial functions correctly; a lack of fluency in determining the expected values and variances of random variables; the failure to set out the logical steps in testing hypotheses in graduation questions.

1 (i) Pr(a life aged 50 dies between exact ages 53 and 58)

(ii) $${}_{3|5}q_{50} = \int_3^8 {}_t p_{50}.\mu_{50+t}\,dt$$

$$= {}_3 p_{50} \int_0^5 {}_t p_{53}.\mu_{53+t}\,dt$$

$$= \exp\left[-\int_0^3 \mu_{50+r}\,dr\right] \int_0^5 \exp\left[-\int_o^t \mu_{53+r}\,dr\right] \mu_{53+t}\,dt$$

Several other alternatives were acceptable, in particular

$$\int_3^8 \exp\left[-\int_0^t \mu_{50+r}\,dr\right] \mu_{50+t}\,dt$$

$$\text{or } \exp\left\{-\int_0^3 \mu_{50+r}\,dr\right\}\left[1-\exp\left\{-\int_0^5 \mu_{53+r}\,dr\right\}\right]$$

$$\text{or } \exp\left\{-\int_0^3 \mu_{50+r}\,dr\right\} - \exp.\left\{-\int_0^8 \mu_{x+r}^{50}\,dr\right\}$$

$$\text{or } \exp\left\{-\int_0^3 \mu_{50+r}.dr\right\} \int_0^5 \exp.\left\{-\int_0^r \mu_{53+s}.ds\right\} \mu_{53+r}.dr$$

(iii) $$\frac{l_{53}-l_{58}}{l_{50}} = \frac{32{,}143.546 - 30{,}795.116}{32{,}669.855} = 0.04127$$

2 Present value of the loss = $\lambda = 20000v^{K+1} - 520\ddot{a}_{\overline{K+1|}}$ where K = curtate future lifetime, of the life aged 50.

(i) EPV loss
$= \mathrm{E}(\lambda) = 20{,}000A_{50} - 520\ddot{a}_{50} = 20{,}000 \times 0.38450 - 520 \times 16.003 = -631.56$

(ii) $$\lambda = 20000v^{K+1} - 520\left(\frac{1-v^{K+1}}{d_{0.04}}\right)$$

$= \text{constant } + \left(520/d_{0.04} + 20{,}000\right)v^{K+1}$

$Var\left(v^{K+1}\right) = {}^2A_{50} - A_{50}^2$ (where ${}^2A$ is @ 8.16% and A @ 4%)

$= {}^2A_{50} - 0.38450^2$,

Then $Var(\lambda) = 33{,}520^2\left({}^2A_{50} - 0.38450^2\right)$

So b = 33520^2 = 1,123,590,400

$$C = -(33{,}520 \times 0.38450)^2 = -166{,}111{,}886$$

An approach using co-variances is possible, but much more complicated.

3 In both cases the retrospective policy value is $\dfrac{D_x}{D_{x+t}}\left(P\ddot{a}_{x:\overline{t}|} - A^{\;1}_{x:\overline{t}|}\right)$.

${}_tV_x$ will exceed ${}_tV_{x:\overline{n}|}$ if $\qquad P_x > P_{x:\overline{n}|}$

$$\frac{1}{\ddot{a}_x} - d > \frac{1}{\ddot{a}_{x:\overline{n}|}} - d$$

$$\ddot{a}_x < \ddot{a}_{x:\overline{n}|}$$

But $\ddot{a}_x = \ddot{a}_{x:\overline{n}|}$ + a nonnegative term.

Hence it is not possible.

Alternatively,

If $\quad {}_tV_x > {}_tV_{x:\overline{n}|}$ then

$$1 - \frac{\ddot{a}_{x+t}}{\ddot{a}_x} > 1 - \frac{\ddot{a}_{x+t:\overline{n-t}|}}{\ddot{a}_{x:\overline{n}|}}$$

which means

$$1 - \frac{\ddot{a}_{x+t}}{\ddot{a}_x} < \frac{\ddot{a}_{x+t:\overline{n-t}|}}{\ddot{a}_{x:\overline{n}|}}$$

But $\ddot{a}_x = \ddot{a}_{x:\overline{n}|} + v^n_n P_x\, \ddot{a}_{x+n}$

$$\ddot{a}_{x+t} = \ddot{a}_{x+t:\overline{n-1}|} + v^{n-t}\;{}_{n-t}P_{x+t}\, \ddot{a}_{x+n}$$

So the inequality will hold if

$$\left(\ddot{a}_{x+t:\overline{n-t}|} + v^{n-t}{}_{n-t}P_{x+t}\, \ddot{a}_{x+n}\right)\ddot{a}_{x:\overline{n}|} < \ddot{a}_{x+t:\overline{n-t}|}\left(\ddot{a}_{x:\overline{n}|} + v^n\;{}_nP_x\, \ddot{a}_{x+n}\right)$$

which means

$$v^{n-t}{}_{n-t}P_{x+t}\, \ddot{a}_{x+n}\, \ddot{a}_{x:\overline{n}|} < \ddot{a}_{x+t:\overline{n-t}|}\, v^n\;{}_nP_x\, \ddot{a}_{x+n}$$

$$\ddot{a}_{x:\overline{n}|} < v^t\;{}_tP_x\, \ddot{a}_{x+t:\overline{n-t}|}$$

$$= \ddot{a}_{x:\overline{n}|} - \ddot{a}_{x:\overline{t}|}$$

which is not true, hence the original assumption is untrue.

[Or: Write in terms of commutation functions

$$\frac{N_{x+t}}{N_x} < \frac{N_{x+t} - N_{x+n}}{N_x - N_{x+n}}$$

which implies

$N_{x+t} > N_x$

Which is not true, hence the original assumption is untrue.

Credit was also given for labelled sketches on ${}_tV_x$ *and* ${}_tV_{x:\overline{n}|}$ *against t, for well chosen numerical examples and for careful verbal arguments.*

It was appreciated that the wording of the question would have been improved if it specifically asked for a comparison for 0≤t≤n.

4 (i) ${}_{0.25}p_{75} = 1 - {}_{0.25}q_{75} = 1 - 0.25 \times q_{75} = 0.98443$ assuming uniform distribution of deaths

$= 1 - 0.25 \times 0.06229 = 1 - 0.01557 = 0.98443$

$p_{75} = {}_{0.75}p_{75} \times {}_{0.25}p_{75.75} = (1 - 0.75 \times q_{75})\,{}_{0.25}p_{75.75}$ assuming uniform distribution of deaths

$$\Rightarrow \quad {}_{0.25}p_{75.75} = \frac{1 - q_{75}}{1 - 0.75q_{75}}$$

$$= \frac{1 - 0.06229}{1 - 0.75 \times 0.06229} = \frac{0.93771}{0.95328} = 0.98366$$

(ii) μ satisfies $e^{-\mu} = p_{75} = 0.93771$ assuming a constant force of mortality

So μ = 0.0643145

Then ${}_{0.25}p_{75} = e^{-0.25\mu} = e^{-0.016078} = 0.98405$

And ${}_{0.25}p_{75.75} = e^{-0.25\mu} = e^{-0.016078} = 0.98405$

Can also use $e^{-0.25\mu} = (0.93771)^{0.25} = 0.98405$

A complete verbal argument that the calendar values of the hazard ratios resulted in the relationships between the survival functions is also acceptable.

5 (i) $L = \exp\left(-512(\sigma+\mu)\right)\exp\left(-20(\rho+\upsilon)\right)\sigma^{10}\rho^{7}\mu^{2}\upsilon^{3}$

(ii) $l = \ln L = -512\sigma + 10\ln\sigma + const.(w.r.t.\ \sigma)$

$$0 = \frac{\partial l}{\partial\sigma} = -512 + \frac{10}{\sigma} \Rightarrow \hat{\sigma} = \frac{10}{512} = 0.0195 \text{ p.a.}$$

A derivation was required.

(iii) $-\dfrac{\partial^2 l}{\partial\sigma^2} = \dfrac{10}{\sigma^2}$, hence we can estimate $Var\,\hat{\sigma}$ by

$\left(\dfrac{10}{\sigma^2}\right)^{-1}$, hence by $\dfrac{\hat{\sigma}^2}{10}$ $\left(= \dfrac{\hat{\sigma}}{512}\right)$

Hence estimated standard deviation of $\hat{\sigma}$ is $\hat{\sigma}/\sqrt{10} = 0.00618$.

6 (i) male non-smokers aged 30 at entry.

(ii) $\dfrac{h_j(t)}{h_i(t)} = \dfrac{\exp(0+0.2+0)}{\exp(0.1+0.2-0.05)} = e^{-0.05} = 0.9512$

where life *j* is a male smoker aged 30 at entry, and
life *i* is a female smoker aged 40 at entry.

But $S(t) = \exp\left(-\int_0^t h(s)ds\right)$, hence

$S_j(t) = \left(S_i(t)\right)^{0.9512}$, which implies that

$S_j(t) > S_i(t)$ for all $t > 0$.

(iii) $\dfrac{h_j(t)}{h_i(t)} = \dfrac{\exp(0+0.2-0.05)}{\exp(0.1+0+0)} = e^{0.05} = 1.0513$,

where life *j* is a female smoker aged 30 at entry, and
life *i* is a male non-smoker aged 40 at entry.

Hence $S_j(t) = \left(S_i(t)\right)^{1.0513}$, which implies that

$S_j(t) < S_i(t)$ for all $t > 0$.

7 (i) EPV $= \displaystyle\int_0^5 2000e^{-0.06s}e^{-0.01s}ds + \int_5^{20} 10000e^{-0.07s}ds$

$= 2000\bar{a}_{\overline{5|}}$ (at a force of 0.07) $+10000e^{-0.35}\bar{a}_{\overline{15|}}$ (at force of 0.07)

$= 2000 \times 4.21874158 + 10000 \times e^{-0.35} \times 9.28660358 = 73879.07$

(ii) EPV = 20,000 $\left(1 - \delta\,.\,\bar{a}_{x:\overline{5|}}\right)$

From (i), $\bar{a}_{x:\overline{5|}} = 4.2187$

Hence EPV = 20,000 (1 – 0.06 × 4.2187) = £14,937.56

Other expressions will lead to the same numerical results in (i) and (ii).

8 (i) Policy year rate interval starting (for deaths classified x) on the policy anniversary on which lives were x next birthday.

(iii) The total number of life years for which lives were exposed to the risk of dying while aged x next birthday on the immediately preceding policy anniversary.

or

Let $P_x(t)$ be a census at time t after the start of the period of investigation of those lives aged x next birthday on the policy anniversary immediately prior to t, then central exposed to risk is

$$\int_{t=0}^{t=T} P_x(t).dt$$

where the period of investigation is (0,T).

(iii) $x+f$ = average age of lives half-way through the rate interval
Thus, if birthdays uniformly distributed over the policy year, lives will be on average x-½ at the start of the rate interval and $x+f = x$.

(iv) The assumption of uniform birthdays over the policy year is violated.

Lives now $x - \frac{1}{6}$ on average at the start of the rate interval and hence $\overset{o}{\mu}_x$ estimates $\mu_{x+\frac{1}{3}}$.

9 (i) Done because life assurance companies have data in this form (and hence far easier than working with lives) and because the result is an unbiased estimate of the rate, provided there is no correlation between deaths and the number of policies covering a life. However, the variance of the estimate is increased. Graduation tests etc will therefore need to be adjusted to allow for this.

(ii) Let $\mathbf{D}_i$ be the number of **deaths** among the $\pi_i N$ lives each with i polices, and let $\mathbf{C}_i$ be the number of **claims** among the same lives (i.e. $\boldsymbol{C}_i = i\boldsymbol{D}_i$). We can say that:

$\mathbf{D}_i \sim \text{Binomial}(\pi_i N, q_x)$

because we have independence of **deaths**. Therefore:

$$\begin{aligned}
\mathrm{Var}[\mathbf{C}] &= \mathrm{Var}[\sum_i \mathbf{C}_i\,] \\
&= \mathrm{Var}[\sum_i i\mathbf{D}_i\,] \\
&= \sum_i i^2 \mathrm{Var}[\mathbf{D}_i] \qquad \text{(independence of deaths)} \\
&= \sum_i i^2 \pi_i N q_x (1-q_x)\,.
\end{aligned}$$

If we observed $\sum_i i\pi_i N$ **independent** lives (or policies) we would have:

$$\mathrm{E}[\mathbf{C}] = \sum_i i\pi_i N q_x$$

$$\mathrm{Var}[\mathbf{C}] = \sum_i i\pi_i N q_x (1-q_x)\,.$$

So, the effect of duplicate policies is to increase the variance of the number of claims, in the ratio

$$r = \frac{\sum_i i^2 \pi_i}{\sum_i i\pi_i}$$

10 The Nelson-Aalen estimator is $\hat{\Lambda}(t) = \sum \frac{d_j}{n_j}$, where:

T	n_j	d_j	d_j/n_j	(i) $\hat{\Lambda}(t) = \sum d_j/n_j$	(ii) $\hat{S}(t) = \exp(-\hat{\Lambda}(t))$
$0 \le t < 30$	12	0	0	0	1
$30 \le t < 35$	12	2	2/12	0.1667	0.8465
$35 \le t < 40$	9	1	1/9	0.2778	0.7575
$40 \le t < 50$	8	2	2/8	0.5278	0.5899
$50 \le t < 68$	6	1	1/6	0.6944	0.4994
$68 \le t < 71$	5	1	1/5	0.8944	0.4088
$71 \le t < 120$	4	1	1/4	1.1444	0.3184

(iii) $\hat{S}(70) = 0.4088$

11 (i) (a) $${}_{t+h}p_x^{00} = \sum_{k=0}^{3} {}_t p_x^{0k}\, {}_h p_{x+t}^{k0} = {}_t p_x^{00}\, {}_h p_{x+t}^{00} = {}_t p_x^{00}\left(1 - \sum_{k=1}^{3} {}_h p_{x+t}^{0k}\right)$$

making Markov assumption, all other transitions have zero probability, by law of total probability

$$= {}_t p_x^{00}\left(1 - \sum_{k=1}^{3} h\mu^{0k} + o(h)\right) \text{ using definitions of } \mu^{ok}\, k = 1,2,3$$

then

$$\Rightarrow \frac{{}_{t+h}p_x^{00} - {}_t p_x^{00}}{h} = -{}_t p_x^{00} \sum_{k=1}^{3} \mu^{0k} + o(h)/h$$

taking limits

hence $\dfrac{\partial}{\partial t}\left({}_t p_x^{00}\right) = \lim_{h \to 0+}\left(\dfrac{{}_{t+h}p_x^{00} - {}_t p_x^{00}}{h}\right) = -\left(\sum_{k=1}^{3} \mu^{0k}\right){}_t p_x^{00}$, and ${}_0 p_x^{00} = 1$.

(b) $${}_{t+h}p_x^{01} = \sum_{k=0}^{3} {}_t p_x^{0k}\, {}_h p_{x+t}^{k1} = {}_t p_x^{00}\, {}_h p_{x+t}^{01} + {}_t p_x^{01}\, {}_h p_{x+t}^{11} = {}_t p_x^{00}(h\mu^{01} + o(h)) + {}_t p_x^{01} \times 1$$

Then

$$\Rightarrow \frac{{}_{t+h}p_x^{01} - {}_t p_x^{01}}{h} = \mu^{01}\, {}_t p_x^{00} + o(h)/h,$$

hence $\dfrac{\partial}{\partial t}\left({}_t p_x^{01}\right) = \mu^{01}\, {}_t p_x^{00}$, and ${}_0 p_x^{01} = 0$.

(ii) (a)

$$\frac{\partial}{\partial t}\left(\ln {}_t p_x^{00}\right) = -\sum_{k=1}^{3} \mu^{0k}$$

then

$$\Rightarrow {}_t p_x^{00} = const \times \exp\left(-t \sum_k \mu^{0k}\right)$$

$${}_0 p_x^{00} = 1 \Rightarrow const = 1.$$

(b)

$$\frac{\partial}{\partial t}\left({}_t p_x^{01}\right) = \mu^{01}\, {}_t p_x^{00} \Rightarrow {}_t p_x^{01} - {}_0 p_x^{01} = \mu^{01} \int_0^t {}_s p_x^{00}\, ds,$$

Now ${}_0 p_x^{01} = 0$ so

$$ {}_t p_x^{01} = \mu^{01}\left(\sum_{k=1}^{3}\mu^{0k}\right)^{-1}\left[-e^{-s\sum \mu^{0k}}\right]_{s=0}^{t} = \frac{\mu^{01}}{\sum_{k=1}^{3}\mu^{0k}}\left(1-e^{-t\sum_{k=1}^{3}\mu^{0k}}\right) $$

12 (i) For all tests we assume H_0 : the graduated rates of mortality are the true underlying mortality rates of the experience.

Chi-squared test: χ^2 = 10.5 with 3 degrees of freedom.

Significant at the 5% level since greater than 7.815.

Candidates who pointed out that the data given was only a sample from the data used to fit the mathematical formula and thus argued that the degrees of freedom should be >3 were also given credit.

NOW any two of

Range	-∞,-3	-3,-2	-2,-1	-1,0	0,+1	+1,+3	+2,+3	+3,∞
Observed	0	1	1	3	2	0	0	0
Expected	0	0.15	0.96	2.39	2.39	0.96	0.15	0

Individual standardized deviations: –2.14 and –1.97 (just) outside ±1.96

2 out of 7 lie outside ±1.960, the upper and lower 2.5% points of standard normal, and both are negative

There are 2 positive and 5 negative deviations, when are equal number of each and are expected if this is true.

There are 3 absolute deviations < 0.67 and 4 greater than 0.67, when an equal number of each expected if H_0 is true.

These results cast some doubt on whether the individual standardized deviations are approximately standard normal, and this whether H_0 is true.

These results cast some doubt on whether the individual standardized deviations are approximately standard normal, and thus whether H_0 is true.

Group signs of deviations (stevens Test)L with n_1=2 positive and n_2=5 negative deviations , these are 2 groups of +ve signs.

Prob (number of groups less than or equal to 2) = $\sum_{t=1}^{2} \frac{\binom{n_1-1}{t-1}\binom{n_2+1}{t}}{\binom{n}{n_1}}$

$= \frac{\binom{1}{0}\binom{6}{1}}{\binom{7}{2}} + \frac{\binom{1}{1}\binom{6}{2}}{\binom{7}{2}}$

$= \frac{6}{21} + \frac{15}{21}$

= 1

Under null hypothesis, the probability of 2 or fewer groups of positive signs is 1 so no reason to reject null hypothesis.

Serial Correlation Test:

$\bar{z} = -\frac{4.07}{7} = -0.5814$

Then

I	1	2	3	4	5	6	7
$z_i - \bar{z}$	-1.56	-0.31	0.15	0.55	1.46	-1.39	1.09
$z_{i+1} - \bar{z}$	-0.31	0.15	0.55	1.46	-1.39	1.09	

And $r_1 = \frac{-2.2219}{6/7 \times 8.1065} = -0.3198$

Then standardised $r_1 = -0.3198\sqrt{7} = -0.8460$

which is standard normal if H_0 is true, so no reason to reject null hypothesis.

Cumulative deviations:

$\frac{169 - 187.03}{\sqrt{186.62}} = -1.32$, which is standard normal if H_0 is true

Not significant at 5% level, or even at 10%.

Signs of deviations: 3 out of 7 positive.

If H_0 is true

$$\frac{7!}{k!(7-k!)}\left(\tfrac{1}{2}\right)^7$$

giving

K	0	1	2	3
Probability	0.00781	0.05469	0.16406	0.27344

So not significant at 5% level.

*Credit was not given for **both** the grouping of signs and serial correlations test*

Comments: The graduated rates are apparently too high over this age range (see individual standardized deviations) but otherwise appear adequate.

There was a misprint in the question. The standardised deviation of –0.43 should be +0.43. The majority of candidates used the standardised deviations as given. Any candidate who recalculated the standardised deviation and used +0.43 in subsequent tests received full credit. An upper case Q in place of the correct lower case q in the column headings caused no confusion to those who noticed it.

(ii) H_0: the graduated rates = true underlying rates

Under H_0, approx. $\sim N(E_x q_x, E_x q_x p_x)$.

If independence can be assumed then $\sum_x (\theta_x - E_x q_x) \sim N(0, \sum E_x q_x p_x)$ approx. and

$$z = \frac{\sum(\theta_x - E_x q_x)}{\sqrt{\sum E_x q_x p_x}} \sim N(0,1) \text{ approx.}$$

Reject H_0 at 5% level of significance if $|z| > 1.96$

But independence assumption is highly dubious. Negative dependence is very likely if one considers the whole age range (because graduation attempts to achieve approximate equality of $\sum \theta_x$ and $\sum E_x q_x$). Can be overcome by splitting the range into two or three equal ranges and testing each separately.

13 (i) $$P\ddot{a}_{45:\overline{20}|} = 10^5\left(\bar{A}^{\ 1}_{45:\overline{20}|} + 2\bar{A}^{\ 1}_{45:\overline{10}|}\right)$$

i.e. $13.488P \approx 10^5 \times 1.04^{½}(0.10434833 + 2\times 0.03641229)$

P = 1,339.57

Using $\bar{A}^{1}_{x:\overline{n}|} = (1+i)^{½}\left(\dfrac{M_x - M_{x+n}}{D_x}\right)$

Other assumptions produce small numerical differences in the answer. All received full credit.

(ii) $${}_{10}V = 10^5\bar{A}^{\ 1}_{55:\overline{10}|} - P\ddot{a}_{45:\overline{10}|} \approx 10^5 \times 1.04^{½} \times 0.10546950 - 8.045P = -21.02$$

Again other assumptions which produce different numerical answers received full credit.

A retrospective calculation is also satisfactory.

(iii) More cover provided in the first 10 years than is paid for by the premiums in those years.

Hence policyholder "in debt" at time 10, size of debt equals negative policy value.

If policy is lapsed during first ten years (possibly longer) the company will suffer a loss

Not possible to recover this loss from policyholder.

(iv) Collect the premiums more quickly

e.g. shorten premium paying term
make premiums larger in earlier years, smaller in later years

Change the pattern of benefits to reduce benefits in first ten years and increase them in last ten years

Other sensible suggestions received credit.