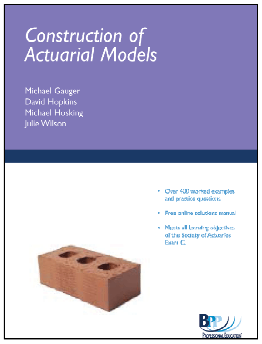

Construction of Actuarial Models

By J. Wilson, M. Hosking, D. Hopkins, and M. Gauger

Published by BPP Professional Education

Solutions to practice questions – Chapter 6

Solution 6.1

There is *left truncation*, at the age at which each life is first observed: *eg* life 8 is left truncated at age $49\frac{4}{12}$.

There is *right censoring*, at the age at which each life is last observed, *except* for those that are observed to die: *eg* life 8 is right censored at age 51, but lives 5, 9 and 10 are *not* censored.

Solution 6.2

The required times are:

$y_2 = 49\frac{1}{12}$ (life 10)

$y_3 = 51\frac{5}{12}$ (life 5)

The required values are:

$s_2 = s_3 = 1$ (lives 10 and 5 respectively).

The required numbers are:

$r_2 = 3$ (because lives 2, 7 and 10 are available to die at age $49\frac{1}{12}$)

$r_3 = 4$ (because lives 1, 3, 4 and 5 are available to die at age $51\frac{5}{12}$).

Solution 6.3

In terms of $S(x)$, we have:

$$\hat{P}\left(X > x \middle| X > 48\tfrac{3}{12}\right) = \frac{\hat{P}(X > x)}{\hat{P}\left(X > 48\frac{3}{12}\right)} = \frac{\hat{S}(x)}{\hat{S}\left(48\frac{3}{12}\right)}$$

In International Actuarial Notation:

$$\hat{P}\left(X > x \middle| X > 48\tfrac{3}{12}\right) = \text{ probability of } \left(48\tfrac{3}{12}\right) \text{ surviving at least } x - 48\tfrac{3}{12} \text{ years}$$

$$= {}_{(x-48.25)}\hat{p}_{48.25}$$

Solution 6.4

The empirical probability of surviving the instant of age $49\frac{1}{12}$ is:

$$\frac{r_2 - s_2}{r_2} = 0.6\dot{6}$$

Because there are no other deaths over $\left[49\frac{1}{12}, 51\frac{5}{12}\right]$ then:

$$\hat{P}\left(X > x \middle| X > 48\tfrac{9}{12}\right) = 0.6\dot{6} \qquad \text{for } \left(49\tfrac{1}{12} \le x < 51\tfrac{5}{12}\right)$$

So, for the same age range $\left(49\frac{1}{12} \le x < 51\frac{5}{12}\right)$:

$$\hat{P}\left(X > x \middle| X > 48\tfrac{3}{12}\right) = 0.6\dot{6} \times \hat{P}\left(X > 48\tfrac{9}{12} \middle| X > 48\tfrac{3}{12}\right) = 0.6\dot{6} \times 0.75 = 0.50$$

Similarly the proportion surviving time $\left(y_3 = 51\frac{5}{12}\right)$ is:

$$\frac{r_3 - s_3}{r_3} = \frac{4-1}{4} = 0.75$$

$$\therefore \qquad \hat{P}\left(X > x \middle| X > 48\tfrac{3}{12}\right) = 0.75 \times 0.5 = 0.375 \qquad \text{for } \left(51\tfrac{5}{12} \le x < 51\tfrac{11}{12}\right)$$

noting that the oldest observed age $\left(y_{\max}\right)$ is $51\frac{11}{12}$. This means we have no information about the survival function beyond this age.

We can now write down the estimate of the survival function over the whole age range as follows:

x	$\hat{P}\left(X > x \middle\| X > 48\frac{3}{12}\right) = \dfrac{\hat{S}(x)}{\hat{S}\left(48\frac{3}{12}\right)}$
$\left(48\frac{3}{12} \le x < 48\frac{9}{12}\right)$	1.000
$\left(48\frac{9}{12} \le x < 49\frac{1}{12}\right)$	0.750
$\left(49\frac{1}{12} \le x < 51\frac{5}{12}\right)$	0.500
$\left(51\frac{5}{12} \le x < 51\frac{11}{12}\right)$	0.375

Notice that the correct use of $\le$ and $<$ are important when stating the applicable age ranges for x.

Solution 6.5

First we express the data in terms of the duration, in months, since the operation, as shown below.

Life	Time of first observation	Time of last observation	Reason for ceasing observation
1	12.5	17	Death
2	10.5	42	Death
3	6.5	24	Censored
4	0	34.5	Censored
5	0	21	Death
6	0	20.5	Censored
7	0	13.5	Censored
8	0	7.5	Censored
9	0	1	Death
10	0	2.5	Censored

The time of first observation is zero (if the life undergoes the operation during the investigation period), or the (average) duration since operation at 1 January 1998 if the operation had taken place before then. The time of last observation is the (average) time since the operation when the individual either died or was censored.

Next we identify:

$y_{\min} = 0$ (the shortest observed time since operation)

$y_{\max} = 42$ (the longest observed time since operation)

Times of death are:

$y_1 = 1$, $y_2 = 17$, $y_3 = 21$, $y_4 = 42$

So now we calculate:

$r_1 =$ number of lives available to die at duration y_1 (*ie* at 1 month)

$= 7$ (lives 4 to 10 inclusive are observed at duration 1 month)

$r_2 =$ number of lives available to die at duration y_2 (*ie* at 17 months)

$= 6$ (lives 1 to 6 inclusive)

$r_3 =$ number of lives available to die at duration y_3 (*ie* at 21 months)

$= 4$ (lives 2 to 5 inclusive)

$r_4 =$ number of lives available to die at duration y_4 (*ie* at 42 months)

$= 1$ (life 2 only)

Solution 6.6

The calculations proceed as follows:

i	y_i	r_i	s_i	$\dfrac{r_i - s_i}{r_i}$	$\prod_{j=1}^{i} \dfrac{r_j - s_j}{r_j}$
1	1	7	1	$6/7$	0.8571
2	17	6	1	$5/6$	0.7143
3	21	4	1	$3/4$	0.5357
4	42	1	1	$0/1$	0

From this we obtain the following estimate of the survival function. (Note that, because $y_{\min}$ is zero, then we can estimate the *unconditional* survival function $S(x)$.)

x	$\hat{S}(x)$
$0 \le x < 1$	1.0000
$1 \le x < 17$	0.8571
$17 \le x < 21$	0.7143
$21 \le x < 42$	0.5357
$x \ge 42$	0

Solution 6.7

We know the survival function is non-zero at ages beyond the last observed age, because (at least) one individual was known to survive to (just) beyond that point. However, we do not know how much longer the individual(s) concerned would have lived for, had they not been censored.

(This is, in fact, precisely the effect of right censoring that we described in Chapter 5.)

Solution 6.8

These two options are biased.

Option (ii) will overstate the value of the survival function at durations beyond $y_{\max}$ (because it assumes that all remaining individuals live for ever beyond this age, which is being over-optimistic for any mortal).

Option (iii) will understate the value of the survival function at durations beyond $y_{\max}$ (because it assumes that all remaining individuals die immediately after they are censored at time $y_{\max}$, which is unduly pessimistic).

Solution 6.9

Now:

$$y_{\max} = 51\tfrac{11}{12} = 51.9167$$

and

$$\hat{S}\left(y_{\max}\right) = 0.375$$

$$\therefore \qquad \mu = \frac{-\ln(0.375)}{51.9167} = 0.01889$$

So:

$$\hat{S}(52) = e^{-52\times 0.01889} = 0.3744$$

$$\hat{S}(55) = e^{-55\times 0.01889} = 0.3538$$

$$\hat{S}(60) = e^{-60\times 0.01889} = 0.3219$$

Solution 6.10

A maximum likelihood estimator (MLE) is one whose realisation (the *estimate*) maximizes the probability of the observed event occurring.

We then know that, because of the properties of any MLE, if we were to increase our sample size without limit the distribution of the estimator would tend to a normal distribution (*ie* the MLE is (asymptotically) unbiased, has (asymptotically) minimum possible variance, and is consistent).

Solution 6.11

We call this a non-parametric estimator because the survival function $\hat{S}(x)$ is *directly* estimated by the estimator, without depending on any intermediate function. A parametric estimator would involve having $S(x)$ defined as a function of one or more other parameters. For example we might have:

$$S(x) = \exp\left[-\int_0^x \left(a + b\,c^t\right) dt\right]$$

so that $S(x)$ is a function of the parameters *a*, *b* and *c*. We could find maximum likelihood estimators of each of these parameters, which would then make:

$$\hat{S}(x) = \exp\left[-\int_0^x \left(\hat{a} + \hat{b}\,\hat{c}^t\right) dt\right]$$

a parametric maximum likelihood estimate of $S(x)$. So, a non-parametric estimator requires no intermediate parameters, whereas a parametric one does!

Solution 6.12

"Unbiased" means that the expected value of the estimator is equal to the true (but unknown) value of the quantity concerned.

A consistent estimator is one whose variance tends to zero as sample size increases. So for a large sample it is highly likely to be very near the target parameter.

Solution 6.13

By removing healthy lives from the population, we will be recording survival rates from a subset of lives whose mortality is higher than average. In other words, the expected value of the estimator will be *smaller* in the presence of this censoring than it would have been without censoring. The expected value is therefore not equal to the true survival function of the whole population under study: *ie* it is biased.

Solution 6.14

The estimate for $S(4)$ is:

$$\hat{S}(4)=0.6$$

The variance of $\hat{S}(4)$ is:

$$\text{var}[\hat{S}(4)]=\frac{S(4)[1-S(4)]}{10}\approx\frac{0.6\times 0.4}{10}=0.024$$

So the confidence interval is (approximately):

$$0.6\pm 1.96\sqrt{0.024}=(0.30,0.90)$$

Solution 6.15

There are 7 lives alive at age 3, 2 of whom survive to age 7. So our point estimate of ${}_4p_3$ is 2/7, and the variance of Y, the number of 3 year olds who survive to age 7, is the variance of a *binomial*$(7,2/7)$ distribution, *ie:*

$$7\left(\frac{5}{7}\right)\left(\frac{2}{7}\right)=\frac{10}{7}$$

So the variance of $\frac{Y}{7}$ is approximately $\frac{10}{7^3}$, and the confidence interval is:

$$\frac{2}{7}\pm 1.96\sqrt{\frac{10}{7^3}}=\frac{2}{7}\pm 0.3347=(-0.049,\ 0.620)$$

Notice that we have a confidence interval that includes values below zero. Of course the value of ${}_4p_3$ cannot possibly be less than zero, so it would be sensible to truncate the confidence interval at zero, to get (0, 0.620). The reason why we get this happening is of course because we are making a number of approximations here, and because the sample sizes are small our confidence interval is fairly wide. We shall see later on how (when we are using the product-moment approach), we can make adjustments to the confidence interval so it does not give us values outside the range (0,1).

Solution 6.16

Since ${}_4q_3 = 1 - {}_4p_3$, the estimates of these functions have the same variances. So the confidence interval has the same width, but it is centered on 5/7 rather than 2/7. So:

$$\frac{5}{7} \pm 0.3347 = (0.38, 1.05)$$

Note that again we have a confidence interval that lies partly outside (0,1).

Solution 6.17

The calculations are performed in the following table:

i	y_i	r_i	s_i	$\frac{s_i}{r_i(r_i - s_i)}$	$\sum_{i=1}^{k} \frac{s_i}{r_i(r_i - s_i)}$
1	1	7	1	0.0238	0.0238
2	17	6	1	0.0333	0.0571
3	21	4	1	0.0833	0.1404
4	42	1	1	–	–

We can now work out the standard errors, noting that the column headed $\sigma_s(x)$ is equal to $\sqrt{\sum_{i=1}^{k} \frac{s_i}{r_i(r_i - s_i)}}$.

x	$\hat{S}(x)$	$\sigma_s(x)$	$\widehat{SE}\left[\tilde{S}(x)\right]$
$0 \le x < 1$	1.0000	0	0
$1 \le x < 17$	0.8571	0.1543	0.1322
$17 \le x < 21$	0.7143	0.2390	0.1707
$21 \le x < 42$	0.5357	0.3747	0.2007
$x \ge 42$	0	–	0

Solution 6.18

The calculations are shown in the following table.

x	$\hat{S}(x)$	$\widehat{SE}\left[\tilde{S}(x)\right]$	Lower limit	Upper limit
$0 \le x < 1$	1.0000	0	1.000	1.000
$1 \le x < 17$	0.8571	0.1322	0.598	1.116
$17 \le x < 21$	0.7143	0.1707	0.380	1.049
$21 \le x < 42$	0.5357	0.2007	0.142	0.929
$x \ge 42$	0	0	0	0

Comment: for two of the age ranges the upper limit of the linear 95% confidence interval exceeds 1.0, whereas in reality this cannot be true. This reflects the fact that the MLE $\tilde{S}(x)$ is only *asymptotically* normally distributed, and for such small samples as involved here the normal distribution is not a good approximation for the actual distribution.

Solution 6.19

The calculations are shown in the following table:

x	$\hat{S}(x)$	$\sigma_S(x)$	U	Lower limit	Upper limit
$0 \le x < 1$	1.0000	0	1	1.000	1.000
$1 \le x < 17$	0.8571	0.1543	0.1407	0.334	0.978
$17 \le x < 21$	0.7143	0.2390	0.2485	0.258	0.920
$21 \le x < 42$	0.5357	0.3747	0.3083	0.132	0.825
$x \ge 42$	0	–	1	0	0

Comment: all intervals fall within the possible range of values of $S(x)$, so this is more realistic than the linear confidence interval calculated previously.

The interval is clearly not symmetric in relation to the point estimate $\hat{S}(x)$, but this more realistically reflects the actual distribution of $\tilde{S}(x)$ for small samples.

Solution 6.20

In this formula:

- P_j is the number of lives under observation at the start of the j th interval
- d_j is the number of lives who are first observed in this interval
- u_j is the number of lives right-censored in this interval
- x_j is the number of lives dying in this interval.

The formula is just saying that the number in the population at the start of any interval is the number in the population at the start of the previous interval, plus the new entrants, less the censored items, less the deaths.

