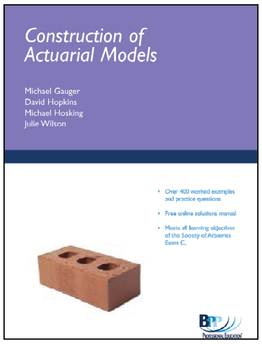


Construction of Actuarial Models

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

Published by BPP Professional Education

Solutions to practice questions – Chapter 7

Solution 7.1

$\mu(x)$ is the annualized probability of dying at exact age x. That is:

$$\mu(x) = \lim_{r \to 0+} \frac{1}{r} \Pr\left(x \le X < x + r \middle| X \ge x\right)$$

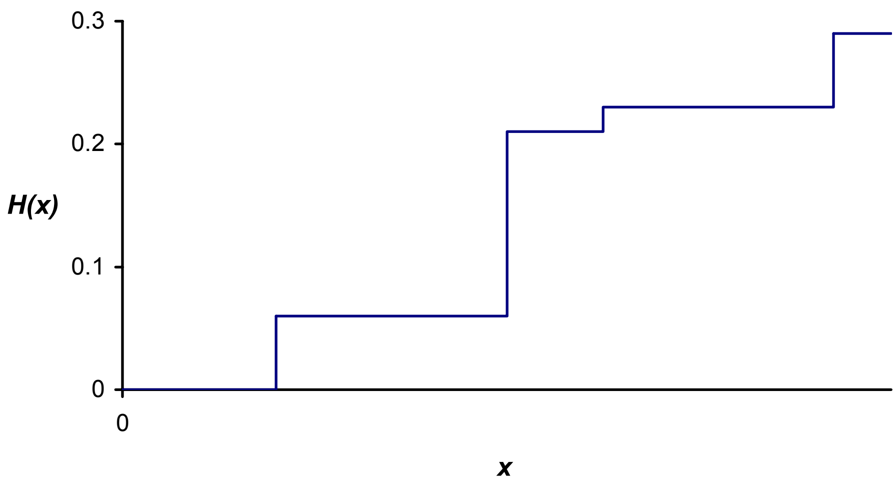
$$= \lim_{r \to 0+} \frac{1}{r} \{\text{Probability that a life who is living at age x, dies before age x+r}\}$$

$$= \lim_{r \to 0+} \frac{{}_r q_x}{r} \quad \text{in International Actuarial Notation}$$

Solution 7.2

$H(x)$ will be a step function, starting at 0 at $x = 0$, increasing by amount $h(x_j)$ at each time $x = x_j$, and remaining flat between these event times.

Graphically the progression might look like this (for example):



Solution 7.3

The only differences between $\hat{H}(x)$ and $H(x)$ are:

- events occur at random times y_i instead of at fixed times x_i
- the observed proportion of lives dying $\left(\frac{s_i}{r_i}\right)$ is used instead of the probability of lives dying $h\left(y_i\right)$.

Intuitively, the probability of an event occurring at a given time is equal to the *expected proportion* of lives suffering the event, so it would follow that the expected value of an estimator $\tilde{H}(x)$ would be the sum of the expected proportions of people dying over all occasions, *ie*:

$$E\left[\tilde{H}(x)\right] = \sum_{y_i \le x} h\left(y_i\right) = H(x)$$

(The take home message from this non-rigorous "proof" is that the estimate $\tilde{H}(x)$ is intuitively reasonable!)

Solution 7.4

The calculations are shown in the following table, where the values of y_i, r_i and s_i are taken from Table 6.2 of Chapter 6.

i	y_i	r_i	s_i	$\lambda_i = \frac{s_i}{r_i}$	$\sum_{k=1}^{i} \lambda_k$
1	48.75	4	1	0.250	0.250
2	49.08	3	1	0.333	0.583
3	51.42	4	1	0.250	0.833

So the estimate of the cumulative hazard function is:

x	$\hat{H}(x) - \hat{H}(48.25)$
$(48.25 \le x < 48.75)$	0
$(48.75 \le x < 49.08)$	0.250
$(49.08 \le x < 51.42)$	0.583
$(51.42 \le x < 51.92)$	0.833

noting that $y_{\min} = 48.25$, $y_{\max} = 51.92$, and because $y_{\max}$ is a censoring time, then we do not have any precise estimate of the cumulative hazard function at ages older than this.

Solution 7.5

The calculations proceed as follows:

i	y_i	r_i	s_i	$\frac{\lambda_i}{r_i} = \frac{s_i}{r_i^2}$	$\sigma_H^2(y_i)$	$\sigma_H(y_i)$
1	48.75	4	1	0.0625	0.0625	0.2500
2	49.08	3	1	0.1111	0.1736	0.4167
3	51.42	4	1	0.0625	0.2361	0.4859

Using $z_{0.025} = 1.96$ we get:

Linear pointwise 95% confidence interval

x	$\hat{H}(x) - \hat{H}(48.25)$	**Lower limit**	**Upper limit**
$48.25 \le x < 48.75$	0	-	-
$48.75 \le x < 49.08$	0.250	– 0.240	0.740
$49.08 \le x < 51.42$	0.583	– 0.234	1.400
$51.42 \le x < 51.92$	0.833	– 0.119	1.785

Log-transformed pointwise 95% confidence interval

x	$\hat{H}(x) - \hat{H}(48.25)$	θ	**Lower limit**	**Upper limit**
$48.25 \le x < 48.75$	0	-	-	-
$48.75 \le x < 49.08$	0.250	7.099	0.035	1.775
$49.08 \le x < 51.42$	0.583	4.059	0.144	2.366
$51.42 \le x < 51.92$	0.833	3.137	0.266	2.613

Comments

The linear confidence intervals are narrower than the log-transformed confidence intervals. (In fact, they are much *too* narrow when calculated from small samples such as this.)

The linear confidence intervals include impossible values for $\hat{H}(x) - \hat{H}(48.25)$ (values less than zero), whereas the log-transformed confidence intervals only include possible values. (Note that it is quite acceptable to have values exceeding 1 for the cumulative hazard function.)

Solution 7.6

The calculations are summarized in the following table:

i	y_i	r_i	s_i	$\lambda_i = \dfrac{s_i}{r_i}$	$\displaystyle\sum_{k=1}^{i} \lambda_k$	$\dfrac{\lambda_i}{r_i} = \dfrac{s_i}{r_i^2}$	$\displaystyle\sum_{k=1}^{i} \dfrac{\lambda_k}{r_k}$
1	1	7	1	0.1429	0.1429	0.0204	0.0204
2	17	6	1	0.1667	0.3095	0.0278	0.0482
3	21	4	1	0.2500	0.5595	0.0625	0.1107
4	42	1	1	1	1.5595	1	1.1107

We can then calculate:

$$\hat{H}(x) = \sum_{y_i \le x} \lambda_i$$

$$\sigma_H(x) = \sqrt{\sum_{y_i \le x} \frac{\lambda_i}{r_i}}$$

and $$\theta = \exp\left(\frac{1.96\,\sigma_H(x)}{\hat{H}(x)}\right)$$

These values, and the resulting 95% pointwise confidence intervals, are shown in the next table:

x	$\hat{H}(x)$	$\sigma_H(x)$	θ	Lower limit	Upper limit
$0 \le x < 1$	0	0	–	–	–
$1 \le x < 17$	0.1429	0.1429	7.0993	0.020	1.014
$17 \le x < 21$	0.3095	0.2195	4.0149	0.077	1.243
$21 \le x < 42$	0.5595	0.3327	3.2073	0.174	1.795
$x \ge 42$	1.5595	1.0539	3.7604	0.415	5.864

Note that, in this case, $y_{\min} = 0$, so we are estimating the full (*ie* not left-truncated) cumulative hazard function at age x. Also the longest time of observation ($y_{\max}$) is due to a death, so that we do have a fully defined estimate of $\hat{H}(x)$ for $x > y_{\max}$.

Comments

The very large size of the confidence interval at $x \ge 42$ reflects the fact that the last observation (a death at time 42) was based on a sample of just one individual.

Solution 7.7

The function $h(x)$ is the *unit-time* hazard rate at exact age x. For example, $\mu(x)$ is the *annual* rate of dying at exact age x. So we need to calculate the rate at which the observed hazards (deaths) are occurring over each unit of time.

It is therefore impossible to estimate the unit-time hazard rate from the proportion dying (λ_i) at a *single* time point y_i. For example, suppose we had observed deaths at ages ($y_i = 65.1, 65.4, 65.5, 65.8$) years. Then we should infer that the annual hazard rate over the year of age (65, 66) was (in the region of) four times the average of the four λ_i values involved.

Solution 7.8

$h(x)$ is, in reality, likely to be changing with age. Provided we can assume that $h(x)$ varies approximately linearly over the age range, then the larger (or smaller) death rates over ages $\left(x,\, x+\frac{1}{2}\right)$ should on average cancel out with the smaller (or larger) death rates over ages $\left(x-\frac{1}{2},\, x\right)$. This means that, on average, the total rate of dying over the whole period of age should be broadly equal to the rate of dying at the central age x.

Solution 7.9

We first need:

$$\hat{h}(20) = \frac{1}{21}\left[\hat{H}(30) - \hat{H}(9)\right] = \frac{1}{21}\sum_{i=3}^{9}\lambda_i$$

$$= \frac{1}{21}(0.0714 + 0.1429 + \cdots + 0.0400) = \frac{0.4990}{21} = 0.0238$$

Next:

$$\hat{h}(21) = \frac{1}{21}\left[\hat{H}(31) - \hat{H}(10)\right] = \frac{1}{21}\sum_{i=3}^{10}\lambda_i$$

$$= \frac{1}{21}(0.0714 + 0.1429 + \cdots + 0.0244) = \frac{0.5234}{21} = 0.0249$$

Continuing:

$$\hat{h}(22) = \frac{1}{21}\left[\hat{H}(32) - \hat{H}(11)\right]$$

but this has the same value as $\hat{h}(21)$, because the same y_i (event times) are included in the two age ranges.

So:

$$\hat{h}(22) = \hat{h}(21) = 0.0249$$

Now:

$$\hat{h}(23) = \frac{1}{21}\left[\hat{H}(33) - \hat{H}(12)\right] = \frac{1}{21}\sum_{i=4}^{10}\lambda_i$$

$$= \frac{1}{21}(0.1429 + 0.0833 + \cdots + 0.0244) = \frac{0.4520}{21} = 0.0215$$

Finally we can also see that:

$$\hat{h}(25) = \hat{h}(24) = \hat{h}(23) = 0.0215$$

because the same event times are included in the age ranges for all three estimates.

Comment

This method produces a step function estimator of $h(x)$, which increases or decreases whenever the age band first includes or excludes a given event time y_i.

Solution 7.10

At $x = 21$ we are using the death data for deaths between ages 10.5 and 31.5 inclusive. This gives us the height of our function from 20.5 to 21.5.

At $x = 22$ we are using the death data for deaths between ages 11.5 and 32.5. This gives us the height of our function from 21.5 to 22.5.

The graph would change if there were any deaths at age $y = 11$ or at age $y = 32$. Deaths at age 11 would be included in the first estimate but not the second, and deaths at ages 32 would be included at the second estimate but not the first.

But there were no deaths at either of these two ages. So the two estimates are the same.

Solution 7.11

At $x = 2$, only the left-hand rectangle makes a contribution. So the height of the rectangle is $\frac{0.2}{6} = 0.0333$.

At $x = 4$, both the left-hand rectangle and the center rectangle will contribute. So the height of the function at this point is $\frac{0.2}{6} + \frac{0.5}{6} = 0.1167$.

At $x = 8$ the points $x = 6$ and $x = 10$ will both contribute. So the height of the curve is $\frac{0.5}{6} + \frac{0.3}{6} = 0.1333$.

We can continue this logic to find the value of the function at each point.

Solution 7.12

The area will be 1. We can check this as follows:

$$4(.0333)+2(.1167)+2(.0833)+2(.1333)+4(.05)=1$$

Solution 7.13

The point 9.6 lies inside the interval from 9½ to 10½. So the value of the smoothed function at this point is:

$$\hat{h}(9.6)=0(0.2)+0(0.5)+1(0.3)=0.3$$

Solution 7.14

Now $k_y(x)=\frac{1}{4}$ for all values for which it is non-zero. So we have:

$$\hat{h}(5)=0.2k_2(5)+0.5k_6(5)+0.3k_{10}(5)=0.2(0)+0.5(1/4)+0.3(0)=0.125$$

$$\hat{h}(9)=0.2k_2(9)+0.5k_6(9)+0.3k_{10}(9)=0.2(0)+0.5(0)+0.3(1/4)=0.075$$

Solution 7.15

Now we know that $k_y(x)=\frac{1}{6}$. So:

$$\hat{h}(8)=0.2k_2(8)+0.5k_6(8)+0.3k_{10}(8)=0.2(0)+0.5(1/6)+0.3(1/6)=0.1333$$

Again, you should be able to see that we can confirm algebraically the values we obtained earlier using our geometric approach.

Solution 7.16

A triangle of height 1 based on (5½, 6½) and a triangle of height 0.6 based on (9½, 10½).

Solution 7.17

The contribution from the left-hand triangle is the function $\hat{h}_1(x) = 0.1111 - 0.0222x$. The contribution from the middle triangle is $\hat{h}_2(x) = 0.0555x - 0.1667$. So the total value of the smoothed function over this interval is $\hat{h}(x) = -0.05556 + 0.0333x$.

Solution 7.18

$$\begin{aligned} f(6.2) &= 0.2k_2(6.2) + 0.5k_6(6.2) + 0.3k_{10}(6.2) \\ &= 0 + 0.5\left(\frac{6 + 0.5 - 6.2}{0.25}\right) + 0 = 0.6 \end{aligned}$$

Solution 7.19

We can recall that the variance of $\tilde{H}(x)$ is:

$$\sigma_H^2(x) = \sum_{y_i \le x} \frac{s_i}{r_i^2} = \sum_{y_i \le x} \frac{\lambda_i}{r_i}$$

So: $$\sigma_H^2(y_i) = \sum_{k=1}^{i} \frac{\lambda_k}{r_k}$$

and $$\sigma_H^2(y_{i-1}) = \sum_{k=1}^{i-1} \frac{\lambda_k}{r_k}$$

$\therefore$ $$\sigma_H^2(y_i) - \sigma_H^2(y_{i-1}) = \frac{\lambda_i}{r_i}$$

So:

$$\sigma_h^2(x) = \sum_{x-b \le y_i \le x+b} \{k_{y_i}(x)\}^2 \left\{\sigma_H^2(y_i) - \sigma_H^2(y_{i-1})\right\}$$

Solution 7.20

First, the formula for the 95% linear confidence interval for $h(x)$ is:

$$\hat{I}(x) = \left\{\hat{h}(x) - 1.96\,\sigma_h(x),\ \hat{h}(x) + 1.96\,\sigma_h(x)\right\}$$

For $h(20)$

$$\hat{h}(20) = 0.0238 \quad \text{(see Example 7.2)}$$

Now:

$$\sigma_h^2(20) = \frac{1}{10.5^2}\sum_{i=3}^{9}\left(\frac{1}{2}\right)^2\frac{\lambda_i}{r_i}$$

$$= \frac{1}{10.5^2}\left\{\frac{1}{4}\left[\frac{0.0714}{14} + \frac{0.1429}{21} + \cdots + \frac{0.04}{25}\right]\right\}$$

$$= \frac{1}{10.5^2}\left\{\frac{0.023959}{4}\right\} = 5.43288 \times 10^{-5}$$

$$\therefore \qquad \sigma_h(20) = \sqrt{5.43288 \times 10^{-5}} = 0.0737$$

So the 95% confidence interval becomes:

$$\left\{0.0238 - 1.96 \times 0.00737,\ 0.0238 + 1.96 \times 0.00737\right\}$$

$$= \left\{0.0094,\ 0.0383\right\}$$