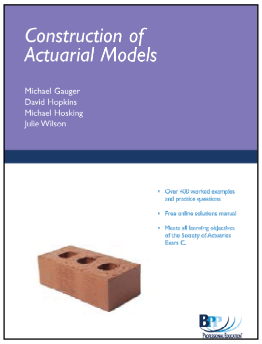

Construction of Actuarial Models

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

Published by BPP Professional Education

Solutions to practice questions – Chapter 8

Solution 8.1

Females who consume no alcohol will experience the baseline hazard rate

---

Solution 8.2

Male 75, 25 units of alcohol

This rate is:

$$h_1(75) = 0.00001 \times 1.1^{75} \exp[0.7 \times 1 + 0.11 \times 25]$$

$$= 0.01272\, e^{3.45}$$

$$= 0.40065$$

Solution 8.3

It's an example of a linear multiple regression model with p independent variables.

Solution 8.4

The hazard rate will increase by an extra factor of:

$$\exp[0.15 \times 1 \times 25] = e^{3.75} = 42.521$$

According to this model, the combination of being male and consuming alcohol is particularly harmful!

Solution 8.5

American: $h(x) = h_0(x)\, e^{\beta_1}$

European: $h(x) = h_0(x)\, e^{\beta_2}$

Australian: $h(x) = h_0(x)\, e^{0} = h_0(x)$

Solution 8.6

Here we would have the models predicting:

$$h\left(x \middle| \text{American}\right) = e^{\beta}$$

$$h\left(x \middle| \text{Australian}\right) = e^{0} \qquad \text{(basically as before).}$$

But: $h\left(x \middle| \text{European}\right) = e^{2\beta}$

This could only be appropriate if the two relative risks are equal:

$$\frac{h\left(x \middle| \text{European}\right)}{h\left(x \middle| \text{American}\right)} = \frac{h\left(x \middle| \text{American}\right)}{h\left(x \middle| \text{Australian}\right)} = e^{\beta}$$

which is very unlikely (and increasingly so the more categories above three you include in the variable).

The problem with having the third covariate Z_3 is that it is not independent of the other covariates. This complicates considerably the tests of significance that we might like to apply when constructing models. In fact, Z_3 would be fully dependent (*ie* completely specified) by the values of the other two variables, and so is completely redundant (as we showed in the previous question), as:

$$Z_3 = \begin{cases} 0 & \text{if } Z_1 = 1 \text{ or } Z_2 = 1 \\ 1 & \text{if } Z_1 = Z_2 = 0 \end{cases}$$

Solution 8.7

At age $y_2 = 75\tfrac{13}{24}$, when life 1 dies, lives 1, 2, 3, 4, 6, 7, 8, 9, 11 and 12 are exposed to risk (*ie* six males and four females). The contribution to the partial likelihood is therefore:

$$\frac{K_1}{\sum_{i \in R(y_2)} K_i} = \frac{e^{\beta}}{6e^{\beta} + 4e^{0}}$$

At age $y_3 = 75\tfrac{21}{24}$, when life 4 dies, lives 3, 4, 8, 9, 11 and 12 are exposed to risk (*ie* two males and four females). The contribution to the partial likelihood is therefore:

$$\frac{K_4}{\sum_{i \in R(y_3)} K_i} = \frac{e^{\beta}}{2e^{\beta} + 4e^{0}}$$

Solution 8.8

The partial likelihood will be:

$$L = \left(\frac{e^0}{5e^\beta + 10e^0} \right) \left(\frac{e^\beta}{5e^\beta + 10e^0} \right) \left(\frac{e^\beta}{5e^\beta + 10e^0} \right) = \frac{e^{2\beta}}{\left(5e^\beta + 10\right)^3}$$

The log-likelihood is:

$$LL = 2\beta - 3\ln\left(5e^\beta + 10\right)$$

Differentiating:

$$\frac{\partial}{\partial\beta} LL = 2 - 3\frac{5e^\beta}{5e^\beta + 10} = 0 \qquad \text{at maximum}$$

Solving:

$$2\left(5e^\beta + 10\right) = 15e^\beta$$

$\therefore \qquad e^\beta = 4$

$\therefore \qquad \hat{\beta} = \ln(4) = 1.3863$

Comment

The fitted model is predicting a relative risk of male mortality of four times that of female mortality. This is what the data have indicated: twice as many males as females have been observed to die out of a population in which the male population is half that of the female population. The "best" (most likely) mortality model to explain this observation is one in which male mortality rates (at any age) are four times the female rates.

Solution 8.9

We have observed twice as many male deaths as female deaths out of a population with equal numbers of the two sexes. The conclusion must be that males are twice as likely to die as females.

So: $\quad e^\beta = 2 \qquad \therefore \ \hat{\beta} = 0.6931$

The sex ratios of the deaths and exposed to risk are now the same. The best explanation of this result is that the mortality rates of the two sexes are the same.

So: $\quad e^\beta = 1 \qquad \therefore \ \hat{\beta} = 0$

Solution 8.10

The remaining two probabilities are:

$$\Pr(\text{1 man and 2 women die}) = \left[h_0(x)e^{\beta}\right]\left[h_0(x)e^{0}\right]^2 50\binom{70}{2}$$

$$\Pr(\text{3 women die}) = \left[h_0(x)e^{0}\right]^3 \binom{70}{3}$$

So the relative risk is:

$$\frac{[h_0(x)]^3 e^{2\beta}}{[h_0(x)]^3\left[e^{3\beta}\binom{50}{3} + e^{2\beta}\binom{50}{2}70 + e^{\beta}50\binom{70}{2} + \binom{70}{3}\right]}$$

$$= \frac{e^{2\beta}}{\left[e^{3\beta}\binom{50}{3} + e^{2\beta}\binom{50}{2}70 + e^{\beta}50\binom{70}{2} + \binom{70}{3}\right]}$$

Solution 8.11

The partial likelihood is:

$$\frac{e^{2\beta}}{\left(50e^{\beta} + 70\right)^3}$$

Solution 8.12

This means that covariate *k* has no effect on mortality and can be omitted from the model.

If the value of the *k*th parameter is not significantly different from zero, then the covariate Z_k is unlikely to be useful (consistent) in predicting the future relative risk. So we can remove it from the model, and end up with a simpler model.

If H_0 is rejected (so β_k is significantly different from zero), then Z_k appears to have some consistent predictive power and should therefore be included in our model.

A useful purpose (*ie* whether or not a covariate should be included in the relative risk model) is therefore served by choosing this particular form of H_0.

Solution 8.13

We will need the value of the log-likelihood (LL) separately using the fitted parameters of the full model, and assuming that both parameters are equal to zero.

Start with the partial likelihood. This is:

$$L = \frac{e^{3(\beta_1+\beta_2)}\, e^{3\beta_1}\, e^{3\beta_2}}{\left(1{,}000\, e^{(\beta_1+\beta_2)} + 2{,}000\, e^{\beta_1} + 3{,}000\, e^{\beta_2} + 4{,}000\right)^9}$$

The log-likelihood is:

$$LL = 6\beta_1 + 6\beta_2 - 9\ln X$$

where:

$$X = 1{,}000\, e^{(\beta_1+\beta_2)} + 2{,}000\, e^{\beta_1} + 3{,}000\, e^{\beta_2} + 4{,}000$$

The first partial derivatives are:

$$\frac{\partial LL}{\partial \beta_1} = 6 - \frac{9\left(1{,}000\, e^{(\beta_1+\beta_2)} + 2{,}000\, e^{\beta_1}\right)}{X}$$

$$\frac{\partial LL}{\partial \beta_2} = 6 - \frac{9\left(1{,}000\, e^{(\beta_1+\beta_2)} + 3{,}000\, e^{\beta_2}\right)}{X}$$

To get the full model fitted parameter values, set to zero and solve for β_1 and β_2 to give:

$$\hat{\beta}_1 = 1.663$$

$$\hat{\beta}_2 = 1.2575$$

So:

$$\hat{\beta}^t = (1.663,\ 1.2575)$$

$$\hat{\beta}^{*t} = (0,\ 0)$$

Using $\hat{\beta}$ we obtain:

$$LL = -78.6328$$

and using $\hat{\beta}^*$ we get:

$$LL^* = -82.8931$$

$\therefore \qquad X^2 = 2(LL - LL^*) = 8.5206$

This is tested against a χ^2 distribution with 2 degrees of freedom, and H_0 is rejected at the 2½% level, but not at the 1% level, because from the Tables:

$$\Pr\left(\tilde{\chi}^2_{[2]} > 7.38\right) = 0.025$$

and $\qquad \Pr\left(\tilde{\chi}^2_{[2]} > 9.21\right) = 0.01$.

Solution 8.14

Here we need to test separately the hypotheses:

$H_0: \quad \beta_1 = 0$

$H_0: \quad \beta_2 = 0$

First test $H_0: \quad \beta_1 = 0$.

Fitting the model with $\beta_1 = 0$ leads to:

$$LL = 6\beta_2 - 9\ln\left(4{,}000e^{\beta_2} + 6{,}000\right)$$

$$\frac{\partial LL}{\partial \beta_2} = 6 - \frac{9 \times 4{,}000e^{\beta_2}}{4{,}000e^{\beta_2} + 6{,}000}$$

which leads to $\hat{\beta}_2^* = 1.09861$.

The log-likelihood evaluated at $\hat{\beta}^* = (0,\ 1.09861)$ is:

$$LL^* = -81.5915$$

$$\therefore \qquad X^2 = 2(LL - LL^*) = 2(-78.6328 - (-81.5915)) = 5.917$$

This is tested against a χ^2 distribution with 1 degree of freedom, and H_0 is rejected at the 2½% level, but not at the 1% level, because from the Tables:

$$\Pr\left(\tilde{\chi}^2_{[1]} > 5.02\right) = 0.025$$

and

$$\Pr\left(\tilde{\chi}^2_{[1]} > 6.64\right) = 0.01 .$$

Now we find the corresponding test results for β_2.

First we test $H_0: \quad \beta_2 = 0$.

Fitting the model with $\beta_2 = 0$ leads to:

$$LL = 6\beta_1 - 9\ln\left(3{,}000e^{\beta_1} + 7{,}000\right)$$

$$\frac{\partial LL}{\partial \beta_1} = 6 - \frac{9 \times 3{,}000e^{\beta_1}}{3{,}000e^{\beta_1} + 7{,}000}$$

which leads to $\hat{\beta}_1^* = 1.54044$

Here:

$$LL^* = -80.3278$$

so:

$$X^2 = 2(LL - LL^*) = 2(-78.6328 - (-80.3278)) = 3.390$$

This is not significant at a 5% level, because from the Tables:

$$\Pr\left(\tilde{\chi}^2_{[1]} > 3.84\right) = 0.05$$

so H_0 is acceptable.

The most appropriate model is therefore one in which sex affects mortality, but smoking does not, *ie*:

$$h(x) = h_0(x)\,e^{\beta_1 Z_1} = h_0(x)\,e^{1.54\,Z_1}$$

using the maximum likelihood estimate of β_1 obtained in part (ii).

Solution 8.15

The relative risk is:

$$\frac{\exp\left[2.5\beta_1 + \beta_2\right]}{\exp\left[\beta_1 + 2.5\beta_2\right]} = \exp\left[1.5\beta_1 - 1.5\beta_2\right] = \exp\left[1.5(\beta_1 - \beta_2)\right]$$

We can work out the 95% confidence interval for $1.5(\beta_1 - \beta_2)$, and then exponentiate in order to convert into terms of relative risk.

Now:

$$E\left[1.5(\tilde{\beta}_1 - \tilde{\beta}_2)\right] = 1.5 \times 0.4 = 0.6$$

To obtain the standard error of $1.5(\tilde{\beta}_1 - \tilde{\beta}_2)$, we need:

$$\text{var}\left[1.5(\tilde{\beta}_1 - \tilde{\beta}_2)\right] = 1.5^2\,\text{var}\left[\tilde{\beta}_1 - \tilde{\beta}_2\right]$$

$$= 1.5^2\left[(1)^2\,\text{var}\left(\tilde{\beta}_1\right) + (-1)^2\,\text{var}\left(\tilde{\beta}_2\right) + 2(1)(-1)\,\text{cov}\left(\tilde{\beta}_1, \tilde{\beta}_2\right)\right]$$

$$= 1.5^2\left[\hat{\sigma}_1^2 + \hat{\sigma}_2^2 - 2\hat{\sigma}_{12}\right] = 1.5^2\,(0.25 + 0.13 - 2 \times 0.06) = 0.585$$

$$\therefore \qquad SE\left[1.5(\tilde{\beta}_1 - \tilde{\beta}_2)\right] = \sqrt{0.585} = 0.765$$

So the 95% confidence interval for $1.5(\tilde{\beta}_1 - \tilde{\beta}_2)$ is:

$$(0.6 - 1.96 \times 0.765,\ 0.6 + 1.96 \times 0.765)$$

$$= (-0.899,\ 2.099)$$

Then the 95% confidence interval for the relative risk is:

$$\left(e^{-0.899},\ e^{2.099}\right) = (0.407,\ 8.158)$$

There is therefore no evidence to suggest that these two types of individual do not have the same hazard rates (because the 95% confidence interval of the relative risk includes unity).

Now we are looking at the following relative risk:

$$\frac{\exp\left[0.8\beta_1 + 1.2\beta_2 - \beta_3\right]}{e^0}$$

The point estimate of $\left[0.8\tilde{\beta}_1 + 1.2\tilde{\beta}_2 - \tilde{\beta}_3\right]$ is:

$$E\left[0.8\tilde{\beta}_1 + 1.2\tilde{\beta}_2 - \tilde{\beta}_3\right] = 0.8\times 2 + 1.2\times 1.6 - 0.8 = 2.72$$

The variance of $\left[0.8\tilde{\beta}_1 + 1.2\tilde{\beta}_2 - \tilde{\beta}_3\right]$ is:

$$\text{var}\left[0.8\tilde{\beta}_1 + 1.2\tilde{\beta}_2 - \tilde{\beta}_3\right] = 0.8^2\ \text{var}\left(\tilde{\beta}_1\right) + 1.2^2\ \text{var}\left(\tilde{\beta}_2\right) + (-1)^2\ \text{var}\left(\tilde{\beta}_3\right)$$

$$+2\left[(0.8)(1.2)\,\text{cov}\left(\tilde{\beta}_1, \tilde{\beta}_2\right) + (0.8)(-1)\,\text{cov}\left(\tilde{\beta}_1, \tilde{\beta}_3\right) + (1.2)(-1)\,\text{cov}\left(\tilde{\beta}_2, \tilde{\beta}_3\right)\right]$$

$$= 0.64\times 0.25 + 1.44\times 0.13 + 0.34 + 2\left[0.96\times 0.06 - 0.8\times 0.04 - 1.2\times 0.05\right]$$

$$= 0.6184$$

$$\therefore \qquad SE\left[0.8\tilde{\beta}_1 + 1.2\tilde{\beta}_2 - \tilde{\beta}_3\right] = \sqrt{0.6184} = 0.7864$$

So the 95% confidence interval for $\left[0.8\tilde{\beta}_1 + 1.2\tilde{\beta}_2 - \tilde{\beta}_3\right]$ is:

$$(2.72 - 1.96\times 0.7864,\ 2.72 + 1.96\times 0.7864) = (1.1787,\ 4.2613)$$

Finally the 95% confidence interval of the relative risk is:

$$\left(e^{1.1787},\ e^{4.2613}\right) = (3.25,\ 70.90)$$

There is therefore evidence to predict that the relative risk between these two types of individual exceeds unity (*ie* that type 1 individuals have higher mortality than type 2 individuals).

(If you found the solution to this question a bit of a surprise, go back and remind yourself how to work out the variance of any linear combination of random variables!)

Solution 8.16

The result would show that none of the covariate factors makes any predictable influence on mortality, and the model should simply be:

$$h(x) = h_0(x)$$

that is, all members of the population aged x are subject to the same hazard rate.

Solution 8.17

Now:

$$H_i(x) = \int_0^x h_i(s)\,ds$$

$$= \int_0^x h_0(s) K_i\,ds$$

$$= K_i \int_0^x h_0(s)\,ds$$

$$= K_i\, H_0(x)$$

$$\therefore \quad \hat{S}_i(x) = \exp\left[-\hat{K}_i\, \hat{H}_0(x)\right]$$

$$= \left(e^{-\hat{H}_0(x)}\right)^{\hat{K}_i} = \left(\hat{S}_0(x)\right)^{\hat{K}_i}$$

Solution 8.18

(i) Estimating the baseline cumulative hazard function

We need to calculate:

$$\hat{H}_0(40) - \hat{H}_0(35) = \sum_{j=1}^{9} \frac{s_j}{W_j}$$

Now, the exposed to risk is identical at each time of death, and consists of 3,000 males and 7,000 females. So, for each y_j:

$$W_j = \sum_{i \in R_j} K_i = 3,000\, e^{1.54} + 7,000$$

$$= 20,993.77$$

As $s_j = 1$ (for all j) then:

$$\hat{H}_0(40) - \hat{H}_0(35) = \sum_{j=1}^{9} \frac{1}{20,993.77} = \frac{9}{20,993.77} = 0.0004287$$

(ii) Estimating probabilities

(a) Pr(male (35) survives at least 5 years)

$$= \frac{S_m(40)}{S_m(35)}$$

$$= \exp\left[-e^{\beta_1}\left(H_0(40) - H_0(35)\right)\right]$$

which is estimated to be:

$$\exp\left[-0.0004287\, e^{1.54}\right] = e^{-0.002} = 0.998$$

(b) Pr(female (35) dies in next 5 years)

$$= 1 - \frac{S_f(40)}{S_f(35)}$$

$$= 1 - \exp\left[-e^0\left(H_0(40) - H_0(35)\right)\right]$$

which is estimated to be:

$$1 - e^{-0.0004287} = 0.0004286$$

Solution 8.19

(a) Standard error of the male survival function

We will need:

$$\text{var}\left[\frac{\tilde{S}_m(40)}{\tilde{S}_m(35)}\right] = 0.998^2\left[R_1 + R_2(\text{male})\right]$$

where:

$$R_1 = \sum_{j=1}^{9} \frac{1}{20,993.77^2} = 2.042 \times 10^{-8}$$

and:

$$R_2(\text{male}) = \left[R_3(\text{male})\right]^t \hat{V} R_3(\text{male})$$

Now R_3 is a one-element vector (scalar) as there is only one parameter. Noting that the value of Z_1 for a male is 1, then:

$$R_3(\text{male}) = \sum_{j=1}^{9}\left[\frac{3,000\,e^{1.54}(1) + 7,000(0)}{3,000\,e^{1.54} + 7,000} - 1\right]\left(\frac{1}{20,993.77}\right)$$

$$= \frac{9(-0.3\dot{3})}{20,993.77} = -1.429 \times 10^{-4}$$

$\hat{V}$, the covariance matrix for the parameters, is then simply $\hat{\sigma}_1^2$, the variance of $\tilde{\beta}_1$. This is obtained from the inverse of the information, as usual, *ie*:

$$\hat{\sigma}_1^2 = \left[-\frac{\partial^2 LL}{\partial \beta_1{}^2} \right]$$

evaluated at $\hat{\beta}_1 = 1.54$. Now:

$$\frac{\partial LL}{\partial \beta_1} = 6 - \frac{9(3{,}000\, e^{\beta_1})}{3{,}000\, e^{\beta_1} + 7{,}000}$$

So:
$$\frac{\partial^2 LL}{\partial \beta_1{}^2} = \frac{-9\left[(3{,}000\, e^{\beta_1} + 7{,}000) 3{,}000\, e^{\beta_1} - 3{,}000^2\, e^{2\beta_1} \right]}{(3{,}000\, e^{\beta_1} + 7{,}000)^2}$$

$$= -2$$

$\therefore$ $\hat{\sigma}_1^2 = 0.5$

and:

$$R_2(\text{male}) = (-1.429 \times 10^{-4})^2\, 0.5 = 1.0210 \times 10^{-8}$$

The required variance is then:

$$0.998^2\, (2.042 + 1.021) \times 10^{-8} = 3.0508 \times 10^{-8}$$

Finally the standard error of our estimate is:

$$\sqrt{3.0508 \times 10^{-8}} = 1.747 \times 10^{-4}$$

(b) Standard error of female distribution function

In (b) we estimated the (conditional) distribution function, *ie*:

$$1 - \frac{S_f(40)}{S_f(35)}$$

so its variance will be the same as in (a), but for females. The calculation of R_2 proceeds as follows:

$$R_2(\text{female}) = \left[R_3(\text{female}) \right] \hat{\sigma}_1^2 \left[R_3(\text{female}) \right]$$

where:

$$R_3(\text{female}) = 9\left[\frac{13{,}993.77}{20{,}993.77} - 0 \right] \left(\frac{1}{20{,}993.77} \right)$$

$$= \frac{9 \times 0.6\dot{6}}{20{,}993.77} = 2.858 \times 10^{-4}$$

$\therefore$ $R_2(\text{female}) = (2.858 \times 10^{-4})^2\, 0.5 = 4.084 \times 10^{-8}$

So the required variance is:

$$(1-4.286\times10^{-4})^2\,(2.042+4.084)\times10^{-8}$$

$$=0.999143\times6.126\times10^{-8}=6.1208\times10^{-8}$$

The standard error is, finally:

$$\sqrt{6.1208\times10^{-8}}=2.474\times10^{-4}$$

Solution 8.20

In the **basic model** there will be three beta factors: β_1, β_2 and β_3.

In the **full model** there will be additional beta factors for the interaction terms, namely:

β_{12}, β_{13}, β_{23} and β_{123}

This makes a total of 7 parameters.

The method of coding used here for weight assumes that the log of the effect for obese individuals will be double the effect for overweight individuals. (Medical researchers refer to this as a "dose-related effect" – being obese is considered to be just a stronger version of being overweight.)

