MTH5131 Actuarial Statistics

Coursework 6 — Solutions

Exercise 1. A variable is a type of covariate (e.g. age) whose actual numerical value enters the linear predictor directly, and a factor is a type of covariate (e.g. sex) that takes categorical values.

Exercise 2. When we set the link function $g(\mu) = \ln \mu$ equal to the linear predictor η and then invert to make μ the subject, we get $\mu = e^{\eta}$. This results in positive values only for μ , which is sensible for a Poisson(μ) distribution, where μ is defined to be greater than 0.

Exercise 3.

1. For the Poisson distribution, we have:

$$f(y) = e^{-\mu}\mu^y/y!$$

We wish to write this in the form:

$$f(y) = \exp\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right].$$

Rearranging the Poisson formula:

$$f(y) = \exp\left[\frac{y\ln\mu - \mu}{1} - \ln y!\right].$$

We can see that this has the correct form with:

$$\theta = \ln\mu,\ b(\theta) = \mu = e^{\theta},\ a(\phi) = \phi = 1,\ c(y, \phi) = -\ln y!$$

2. Using the rearranged form for the Poisson distribution from part 1., we see that the $\ln$ of the likelihood function can be written:

$$\ln L(\mu_1, \mu_2, \mu_3) = \sum_i y_i \ln\mu_i - \sum_i \mu_i - \sum_i \ln y_i! \tag{1}$$

This now becomes, for Model A:

$$\begin{aligned} \ln L &= \alpha\sum_{i=1}^{10} y_i + \beta\sum_{i=11}^{15} y_i + \gamma\sum_{i=16}^{35} y_i - 10e^{\alpha} - 5e^{\beta} - 20e^{\gamma} - \sum_{i=1}^{35}\ln y_i! \\ &= 11\alpha + 3\beta + 4\gamma - 10e^{\alpha} - 5e^{\beta} - 20e^{\gamma} - \sum_{i=1}^{35}\ln y_i! \end{aligned} \tag{2}$$

Differentiating this log-likelihood function in turn with respect to α, β, and γ, we get

$$\frac{\partial}{\partial\alpha}\ln L = 11 - 10e^{\alpha}$$

$$\frac{\partial}{\partial\alpha}\ln L = 3 - 5e^{\beta}$$

and

$$\frac{\partial}{\partial\gamma}\ln L = 4 - 20e^{\gamma}$$

Setting each of these expressions equal to zero in turn, we find that

$$\hat{\alpha} = \ln 1.1 = 0.09531$$

$$\hat{\beta} = \ln 0.6 = -0.51083$$

$$\hat{\gamma} = \ln 0.2 = -1.60944$$

These are the maximum likelihood estimates for α, β, and γ.

3. (a) In this case the log-likelihood function reduces to:

$$\ln L = \alpha \sum_{i=1}^{35} y_i - 35e^{\alpha} - \sum_{i=1}^{35} \ln y_i! = 18\alpha - 35e^{\alpha} - \sum_{i=1}^{35} \ln y_i! \tag{3}$$

Differentiating this with respect to α and setting the result equal to zero, we find that

$$18 - 35e^{\alpha} = 0 \Rightarrow \hat{\alpha} = \ln(18/35) = -0.66498$$

(b) The scaled deviance for Model A is given by:

$$\text{Scaled Deviance} = 2(\ln L_S - \ln L_A)$$

where $\ln L_S$ is the value of the log likelihood function for the saturated model, and $\ln L_A$ is the value of the log-likelihood function for Model A.
For the saturated model, we replace the μ_i's with the y_i's in Equation (1). So:

$$\ln L_S = \sum_i y_i \ln y_i - \sum_i y_i - \sum_i \ln y_i!$$

$$= 4 \times 2\ln 2 - 18 - 4\ln 2 = 4\ln 2 - 18 = -15.2274$$

For the log likelihood for Model A, we replace the parameters α, β, and γ with their estimates $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ which were derived from (2):

$$\ln L_A = 11\hat{\alpha} + 3\hat{\beta} + 4\hat{\gamma} - 10e^{\hat{\alpha}} - 5e^{\hat{\beta}} - 20e^{\hat{\gamma}} - \sum_{i=1}^{35} \ln y_i!$$

$$= 11\ln 1.1 + 3\ln 0.6 + 4\ln 0.2 - 11 - 3 - 4 - 4\ln 2 = -27.6944$$

So the scaled deviance is twice the difference in the log likelihoods:

$$\text{Scaled Deviance} = 2(\ln L_S - \ln L_A) = 2((-15.2274) - (-27.6944)) = 24.93$$

as required.
We now repeat the process for Model B.
Using the value of $\hat{\alpha}$ derived from Equation (3), the log likelihood for Model B is

$$\ln L_B = 18\hat{\alpha} - 35e^{\hat{\alpha}} - \sum_{i=1}^{35} \ln y_i!$$

$$= 18\ln(18/35) - 18 - 4\ln 2 = -32.7422$$

The scaled deviance is again twice the difference in the log likelihoods:

$$\text{Scaled Deviance} = 2(\ln L_S - \ln L_B) = 2((-15.2274) - (-32.7422)) = 35.03$$

(c) We can use the chi-squared distribution to compare Model A with Model B. We calculate the difference in the scaled deviances (which is just $2(\ln L_S - \ln L_B)$)

$$35.03 - 24.93 = 10.1$$

This should have a chi-squared distribution with $3 - 1 = 2$ degrees of freedom, which has a critical value at the upper 5% level of 5.991. Our value is significant here, since $10.10 > 5.991$, so this suggests that Model A is a significant improvement over Model B. We prefer Model A here.

Exercise 4. 1. In parameterised form, the linear predictors are (with i, j and k corresponding to the levels of YO , FS and TC respectively):

$$\text{Model 1}: \quad \alpha_i + \beta_j + \gamma_k \quad \text{(4 parameters)}$$

There is one parameter to set the base level for the combination YO_0, FS_0, TC_0, and one additional parameter for each of the higher levels of the three factors.

$$\text{Model 2}: \quad \alpha_{ij} + \gamma_k \quad \text{(5 parameters)}$$

There are four parameters for the 2×2 combinations of YO and FS (assuming TC) and one additional parameter for the higher level of TC.

$$\text{Model 3}: \quad \alpha_{ijk} \quad \text{(8 parameters)}$$

There are eight parameters for the $2 \times 2 \times 2$ combinations of YO, FS and TC.

2. Model 1 does not allow for the possibility that there may be interactions (correlations) between some of the factors. For example, it may be the case that young drivers tend to drive fast cars and to live in towns.

   With Model 3, which is a saturated model, it would be possible to fit the average values for each group exactly ie there are no degrees of freedom left. This defeats the purpose of applying a statistical model, as it would not 'smooth' out any anomalous results.

3. Normal error structure means that the randomness present in the observed values in each category (eg young/fast/town) is assumed to follow a normal distribution.

   The link function is the function applied to the linear estimator to obtain the predicted values. Associated with each type of error structure is a 'canonical' or 'natural' link function. In the case of a normal error structure, the canonical link function is the identity function

4. The completed table, together with the differences in the scaled deviance and degrees of freedom, is shown below.

Model	Scaled Deviance	Degrees of Freedom	Δ Scaled Deviance	Δ Degrees of Freedom
1	50	7		
$YO + FS + TC$	10	4	40	3
$YO + FS + YO.FS + TC$	5	3	5	1
$YO * FS * TC$	0	0	5	3

   Comparing the constant model and Model 1

   The difference in the scaled deviances is 40.

   This is greater than 7.815, the upper 5% point of the χ^2_3 distribution.

   So Model 1 **is** a significant improvement over the constant model.

   Alternatively, if we use the AIC to compare models, we find that since Δ(deviance)$> 2 \times \Delta$ degrees of freedom, because $40 > 2 \times 3$, Model 1 **is** a significant improvement over the constant model.

   Comparing Model 1 and Model 2

   The difference in the scaled deviances is 5.

This is greater than 3.841, the upper 5% point of the χ_1^2 distribution.

So Model 2 **is** a significant improvement over Model 1.

Alternatively, if we use the AIC to compare models, we find that since Δ(deviance)$> 2 \times \Delta$ degrees of freedom, because $5 > 2 \times 1$, Model2 **is** a significant improvement over model 1.

Comparing Model 2 and Model 3

The difference in the scaled deviances is 5.

This is less than 7.815, the upper 5% point of the χ_3^2 distribution.

So Model 3 is **not** a significant improvement over Model 2.

Alternatively, if we use the AIC to compare models, we find that since Δ(deviance)$\not> 2 \times \Delta$ degrees of freedom, because $5 < 2 \times 3$, Model2 is **not** a significant improvement over model 1.