MAST90139 Assignment 3

Haonan Zhong 867492 (Wed 14:15 Qiuyi Li)

Question 1a

The model fitted in the `R` output is a multi-categorical logistic model, which use status as the nominal categorical response variable, and year as the predictor. The response variable status with three levels follows a multinomial distribution M(371, $\boldsymbol{\pi}$), where $\boldsymbol{\pi}$ is a vector that contains probability of being in each level.

The model fitted to the data is

$$log\frac{P(mild)}{P(normal)} = -4.2917 + 0.0836 \times year$$

$$log\frac{P(severe)}{P(normal)} = -5.0598 + 0.1093 \times year$$

Question 1b

0.1093 is the coefficient for year when status = severe, therefore the intepretation is given by: for every additional year spent working at the coal face the odds of a miner's pneu-monoconiosis status is severe will be increased by a factor of $e^{0.1093} = 1.115497$.

$$\frac{P(severe)}{P(normal)} = e^{-5.0598+0.1093\times year}$$

$$Odds\ ratio = \frac{e^{-5.0598+0.1093\times 10}}{e^{-5.0598+0.1093\times 0}} = e^{0.1093\times 10} = 2.98321$$

Therefore, the 95% confidence interval for the odds ratio of severe status versus normal status for every 10 more years spent working at the coal face is given by the following:

$$(e^{10\times(1.093-1.96\times 0.0165)}, e^{10\times(1.093+1.96\times 0.0165)}) = (2.158903, 4.122254)$$

Question 1c

Estimate the pneumonoconiosis status probabilities for a miner who has spent 25 years working at the coal face, using normal level as the pivot.

$$\frac{P(severe)}{P(normal)} = e^{-5.0598+0.1093\times 25}$$

$$\frac{P(mild)}{P(normal)} = e^{-4.2917+0.0836\times 25}$$

Therefore,

$$P(normal) = \frac{1}{1 + e^{-4.2917+0.0836\times 25} + e^{-5.0598+0.1093\times 25}}$$

Hence,

$$P(mild) = P(normal) \times e^{-4.2917+0.0836\times 25}$$

$$P(severe) = P(normal) \times e^{-5.0598+0.1093\times 25}$$

```
odds_mild = exp(-4.2917 + 0.0836*25)
odds_severe = exp(-5.0598 + 0.1093*25)
p_normal = 1/(1 + odds_mild + odds_severe)
p_mild = p_normal * odds_mild
p_severe = p_normal * odds_severe
cbind(p_normal, p_mild, p_severe)
```

```
##       p_normal    p_mild   p_severe
## [1,] 0.8276955 0.0915555 0.08074898
```

Therefore, the probability of normal, mild and severe is 0.8276955, 0.0915555 and 0.08074898 respectively. And given they sum up to one, which means our estimated probabilities are valid.

Question 2a

The model fitted in question 2 is a cumulative logistic model, response variable is status as an ordinal variable and covariate is year. The response variable status follows a multinomial distribution. And the model can be written as follow

$$log\frac{P(Status = Normal)}{P(Status > Normal)} = 3.9558 - 0.0959 \times year$$

$$log\frac{P(Status < Severe)}{P(Status = Severe)} = 4.8690 - 0.0959 \times year$$

Question 2b

The interpretation of the coefficient estimate 0.0959 is given by: the odds of status moving from normal to mild/severe, or moving from normal/mild to severe increases by a factor of $e^{0.0959} = 1.100649$ for every additional year a coal miner spent working at the coal face.

$$\frac{P(Status = Normal)}{P(Status > Normal)} = e^{3.9558-0.0959\times year}$$

$$\frac{P(Status > Normal)}{P(Status = Normal)} = \frac{1}{e^{3.9558-0.0959\times year}}$$

$$Odds\ ratio = \frac{e^{3.9558-0.0959\times 0}}{e^{3.9558-0.0959\times 10}} = \frac{1}{e^{-0.0959\times 10}} = e^{0.0959\times 10} = 2.6091$$

The estimated odds ratio of non-normal status versus normal status for every 10 more years spent working at the coal face is 2.6091.Furthermore, the 95% confidence interval for the odds ratio of severe status versus normal status for every 10 more years spent working at the coal face is given by the following:

$$(e^{10\times(0.0959-1.96\times 0.01194)}, e^{10\times(0.0959+1.96\times 0.01194)}) = (2.064682, 3.297036)$$

Question 2c

Estimate the pneumonoconiosis status probabilities for a miner who has spent 25 years working at the coal face

$$\frac{P(Status > Normal)}{P(Status = Normal)} = \frac{P(Mild) + P(Severe)}{P(Normal)} = \frac{1 - P(Normal)}{P(Normal)} = \frac{1}{e^{3.9558 - 0.0959 \times 25}}$$

Solving the above equation, we have

$$P(Normal) = \frac{1}{e^{3.9558 - 0.0959 \times 25} + 1} = 0.8261$$

$$\frac{P(Status < Severe)}{P(Status = Severe)} = \frac{P(Normal) + P(Mild)}{P(Severe)} = \frac{1 - P(Severe)}{P(Severe)} = \frac{1}{e^{4.8690 - 0.0959 \times 25}}$$

Solving the above equation, we have

$$P(Severe) = \frac{1}{e^{4.8690 - 0.0959 \times 25} + 1} = 0.0779$$

And since probability sum up to one, therefore

$$P(Mild) = 1 - P(Normal) - P(Severe) = 1 - 0.8261 - 0.0779 = 0.0960$$

Therefore, the estimated probabilities are P(Normal) = 0.8261, P(Mild) = 0.0960, and P(Severe) = 0.0779.

Question 3a

Let π_{it} = P(y_{it} = 1 | age, smoke), t = 1,...,4, i = 1,...,537, and the model can be written as follow,

$$log \frac{\hat{\pi}_{it}}{1 - \hat{\pi}_{it}} = -1.880 - 0.113 \times age + 0.265 \times smoke$$

The mean is,

$$mean = \frac{e^{-1.880 - 0.113 \times age + 0.265 \times smoke}}{1 + e^{-1.880 - 0.113 \times age + 0.265 \times smoke}}$$

The variance is,

$$V\hat{A}R(y_{it}) = \frac{0.9985 \times e^{-1.880 - 0.113 \times age + 0.265 \times smoke}}{(1 + e^{-1.880 - 0.113 \times age + 0.265 \times smoke})^2}$$

The correlation coefficient is,

$$C\hat{o}rr(y_{it}, y_{is}) = 0.3543$$

Question 3b

As we can see from the R output, the design matrix for data where `id` = 536 can be written as follow,

$$Z(\texttt{id} = 536) = \begin{bmatrix} 1 & -2 & 1 \\ 1 & -1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Question 3c

The estimated odds ratio of wheezing for a child for every one year older in age can calculated as follow,

$$\hat{OR} = e^{\hat{\beta}_0} = e^{-0.113} = 0.893$$

And the approximate standard error of the odds ratio estimate can be calculated as follow,

$$s.e.(\hat{OR}) = e^{\hat{\beta}_1} \times s.e.(\hat{\beta}_1) = e^{-0.113} \times 0.044 = 0.039$$

Question 3d

The log odds ratio of wheezing for a 10-year old child with a smoking mother,

$$log\frac{\hat{\pi}_{it}}{1-\hat{\pi}_{it}} = -1.880 - 0.113 \times 1 + 0.265 \times 1$$

The log odds ratio of wheezing for a 9-year old child with a non-smoking mother

$$log\frac{\hat{\pi}_{it}}{1-\hat{\pi}_{it}} = -1.880 - 0.113 \times 0 + 0.265 \times 0$$

Thus, the odds ratio can be calculated as follow,

$$\hat{OR} = \frac{e^{-1.880-0.113\times1+0.265\times1}}{e^{-1.880-0.113\times0+0.265\times0}} = e^{-0.113+0.265} = 1.164$$