Coursework 7: STAT 570

Philip Pham

November 24, 2018

1. Create a binary variable Z_i, with $Z_i = 0$ corresponding to $Y_i \in \{0, 1\}$ and $Z_i = 1$ corresponding to $Y_i \in \{2, 3\}$. Let $q(x_i) = \mathbb{P}(Z_i = 1 \mid x_i)$, with $\mathbf{x}_i = \begin{pmatrix} 1 & x_{1i} & x_{2i} \end{pmatrix}^\intercal$, represent the probability of mental impairment being *Moderate* or *Impaired*, given covariates $\mathbf{x}_i$, $i = 1, \ldots, n = 40$. Provide a single plot that shows the association between $q(x_i)$ and x_{1i} and x_{2i}, on a response scale you feel is appropriate. Comment on the plot.

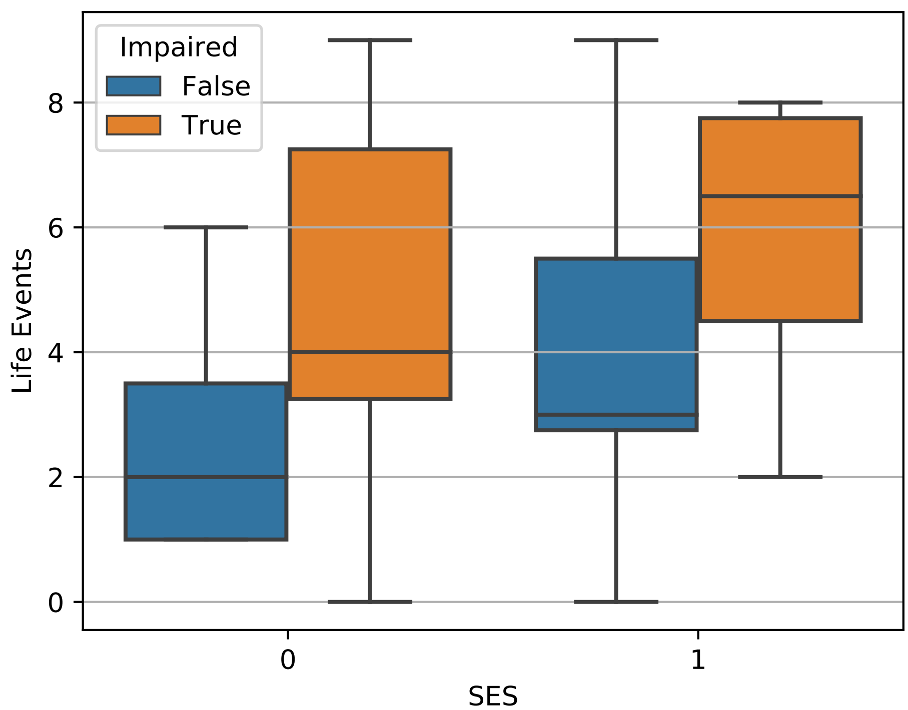

Figure 1: Orange denotes $Z_i = 1$ and blue denotes $Z_i = 0$.

Solution: See Figure 1. Conditioned on SES, those that are impaired ($Z_i = 1$) have a greater number of life events on average.

2. Suppose $Z_i \mid q_i \sim \text{Binomial}(1, q_i)$ independently for $i = 1, \ldots, n = 40$, where $q_i = q(x_i)$. Consider the logistic regression model,

$$q(x_i) = \log\left(\frac{q(\mathbf{x}_i)}{1 - q(\mathbf{x}_i)}\right) = \mathbf{x}_i^\intercal \boldsymbol{\gamma} = \gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i}, \tag{1}$$

where $\boldsymbol{\gamma} = \begin{pmatrix} \gamma_0 & \gamma_1 & \gamma_2 \end{pmatrix}^\intercal$. Write down the log-likelihood $l(\boldsymbol{\gamma})$ for the sample z_i, $i = 1, \ldots, n$.

Solution: Solving for $q(\mathbf{x}_i)$ in Equation 1, we find

$$q(\mathbf{x}_i) = \frac{\exp(\mathbf{x}_i^\intercal \boldsymbol{\gamma})}{1 + \exp(\mathbf{x}_i^\intercal \boldsymbol{\gamma})} = \frac{1}{1 + \exp(-\mathbf{x}_i^\intercal \boldsymbol{\gamma})}. \tag{2}$$

The likelihood function is $L(\boldsymbol{\gamma}) = \prod_{i=1}^n (q(\mathbf{x}_i))^{z_i} (1 - q(\mathbf{x}_i))^{1-z_i}$, so the log-likelihood function becomes

$$\begin{aligned} l(\boldsymbol{\gamma}) = \log L(\boldsymbol{\gamma}) &= \sum_{i=1}^n (z_i \log q(\mathbf{x}_i) + (1 - z_i) \log(1 - q(\mathbf{x}_i))) \qquad (3) \\ &= \sum_{i=1}^n \left(z_i \log \frac{q(\mathbf{x}_i)}{1 - q(\mathbf{x}_i)} + \log(1 - q(\mathbf{x}_i)) \right) \\ &= \sum_{i=1}^n \left(z_i \mathbf{x}_i^\intercal \boldsymbol{\gamma} + \log \frac{1}{1 + \exp(\mathbf{x}_i^\intercal \boldsymbol{\gamma})} \right) = \sum_{i=1}^n -\log(1 + \exp((1 - 2z_i) \mathbf{x}_i^\intercal \boldsymbol{\gamma})). \end{aligned}$$

3. Fit the model described in the previous part, and give confidence intervals for the odds ratios.

   Carefully interpret these odds ratios.

	Estimate	Standard error	95% CI lower bound	95% CI upper bound
γ_0	-0.925065	0.723346	-2.342797	0.492666
γ_1	-1.629731	0.780849	-3.160167	-0.099296
γ_2	0.309899	0.147920	0.019980	0.599818

Table 1: Estimates and confidence intervals for $\hat{\boldsymbol{\gamma}}$ using maximum likelihood estimation.

Solution: Taking the derivative of Equation 3, we have the score function:

$$\begin{aligned} S(\boldsymbol{\gamma}) = \nabla^\intercal l(\boldsymbol{\gamma}) &= \sum_{i=1}^n \frac{2z_i - 1}{1 + \exp((1 - 2z_i) \mathbf{x}_i^\intercal \boldsymbol{\gamma})} \exp((1 - 2z_i) \mathbf{x}_i^\intercal \boldsymbol{\gamma}) \mathbf{x}_i. \\ &= \sum_{i=1}^n \frac{2z_i - 1}{1 + \exp((2z_i - 1) \mathbf{x}_i^\intercal \boldsymbol{\gamma})} \mathbf{x}_i \\ &= X^\intercal (\mathbf{z} - \mathbf{q}(X)), \qquad (4) \end{aligned}$$

where $\mathbf{z} = \begin{pmatrix} z_1 & z_2 & \cdots & z_n \end{pmatrix}^\intercal$ and $\mathbf{q}(X) = \begin{pmatrix} q_1 & q_2 & \cdots & q_n \end{pmatrix}^\intercal$.
From Equation 4, we have the Fisher information matrix:

$$\begin{aligned} I_n(\boldsymbol{\gamma}) &= \operatorname{var}(S(\boldsymbol{\gamma}) \mid \boldsymbol{\gamma}) = \mathbb{E}[S(\boldsymbol{\gamma}) S(\boldsymbol{\gamma})^\intercal \mid \boldsymbol{\gamma}] \\ &= \mathbb{E}[X^\intercal (\mathbf{z} - \mathbf{q}(X)) (\mathbf{z} - \mathbf{q}(X))^\intercal X \mid \boldsymbol{\gamma}] \\ &= X^\intercal \mathbb{E}[(\mathbf{z} - \mathbf{q}(X)) (\mathbf{z} - \mathbf{q}(X))^\intercal \mid \boldsymbol{\gamma}] X \\ &= \sum_{i=1}^n q(\mathbf{x}_i)(1 - q(\mathbf{x}_i)) \mathbf{x}_i \mathbf{x}_i^\intercal = \sum_{i=1}^n \frac{1}{2 + \exp(-\mathbf{x}_i^\intercal \boldsymbol{\gamma}) + \exp(\mathbf{x}_i^\intercal \boldsymbol{\gamma})} \mathbf{x}_i \mathbf{x}_i^\intercal, \qquad (5) \end{aligned}$$

SES	Life Events	Count	Estimate	95% CI lower bound	95% CI upper bound
0	0	1	0.396506	0.096059	1.636675
	1	3	0.540551	0.158334	1.845440
	2	2	0.736926	0.249432	2.177188
	3	3	1.004642	0.368208	2.741129
	4	3	1.369616	0.501460	3.740769
	5	2	1.867180	0.630203	5.532120
	6	1	2.545502	0.742501	8.726699
	8	1	4.730948	0.915136	24.457420
	9	2	6.449640	0.982321	42.346488
1	0	1	0.077708	0.011740	0.514377
	1	2	0.105938	0.020316	0.552412
	2	2	0.144424	0.034495	0.604687
	3	5	0.196892	0.056880	0.681549
	4	2	0.268420	0.089704	0.803195
	5	2	0.365934	0.132703	1.009076
	6	1	0.498873	0.181299	1.372728
	7	2	0.680107	0.228639	2.023045
	8	3	0.927182	0.270204	3.181546
	9	2	1.264015	0.305120	5.236406

Table 2: Estimates for the odds ratios given $\mathbf{x}_i$ with $\hat{\boldsymbol{\gamma}}$.

where we have used independence of the observations and variance of the binomial distribution to get the last line.

We solve Equation 4, $S(\hat{\boldsymbol{\gamma}}) = \mathbf{0}$, to get an estimate for $\boldsymbol{\gamma}$. Using Equation 5, we have that

$$\hat{\boldsymbol{\gamma}} \xrightarrow{\mathcal{D}} \mathcal{N}\left(\gamma, I_n^{-1}(\hat{\gamma})\right), \tag{6}$$

that is, $\hat{\boldsymbol{\gamma}}$ is asymptotically normal.

Using Equation 6, we obtain the estimates and intervals in Table 1.

The predicted log odds ratio given some $\mathbf{x}_i$ is

$$\hat{\theta}_i = \mathbf{x}_i^\intercal \hat{\boldsymbol{\gamma}}, \tag{7}$$

which will have variance

$$\operatorname{var}\left(\hat{\theta}_i\right) = \mathbf{x}_i^\intercal \operatorname{var}(\hat{\gamma})\,\mathbf{x}_i \approx \mathbf{x}_i^\intercal I_n^{-1}(\hat{\gamma})\,\mathbf{x}_i, \tag{8}$$

using Equation 6.

From Equation 8, we can compute confidence intervals for the log odds ratio and exponentiate to get confidence intervals for the odds ratio since log is a monotonic transformation. Doing so results in the estimates in Table 2.

The odds ratio is how much more likely one is to have `Moderate` or `Impaired` mental impairment. Exponentiating Equation 7, we have

$$\exp(\theta_i) = \exp(\gamma_0)\exp(\gamma_1 x_{1i})\exp(\gamma_2 x_{2i}). \tag{9}$$

$\exp(\gamma_0)$ is the expected odds ratio for a subject with 0 SES and no life events. $\exp(\gamma_1)$ is the expected odds ratio between a subject with SES 1 and SES 0. $\exp(\gamma_2)$ is the expected odds ratio for a subject with an additional life event.

4. We will now consider analyses that do not coarsen the data. We begin by defining notation in a generic situation. Suppose the random variable, Y_i, for individual i, $i = 1, \ldots, n$, can take values $0, 1, 2, \ldots, J-1$ (so that that there are J levels). Assume that for individual i, the data follow a multinomial distribution, $Y_i \mid p_i \sim \text{Multinomial}\left(1, \mathbf{p}_i\right)$ independently, where $p_i = \begin{pmatrix} p_{i0} & \cdots & p_{i,J\text{âĹŠ}1} \end{pmatrix}^\intercal$, and p_{ij} represents the probability

$$p_{ij} = \mathbb{P}\left(Y_i = j \mid \mathbf{x}_i\right), \text{ for } j = 0, 1, \ldots, J-1, \tag{10}$$

where $\mathbf{x}_i = \begin{pmatrix} 1 & x_{1i} & x_{2i} & \cdots & x_{ki} \end{pmatrix}^\intercal$ for $i = 1, \ldots, n$.

Suppose the response categories re nominal, that is, have no ordering. In this case, we may consider the *generalized logit model*:

$$p_{ij} = \frac{\exp\left(\mathbf{x}_i^\intercal \boldsymbol{\beta}_j\right)}{\sum_{l=0}^{J-1} \exp\left(\mathbf{x}_i^\intercal \boldsymbol{\beta}_l\right)}, \text{ for } j = 0, \ldots, J-1, \tag{11}$$

where $\boldsymbol{\beta}_j = \begin{pmatrix} \beta_{j0} & \beta_{j1} & \cdots & \beta_{jk} \end{pmatrix}^\intercal$.

Identifiability may be enforced by taking $\boldsymbol{\beta}_{J-1} = \mathbf{0}$, to give

$$\log \frac{p_{ij}}{p_{i,J-1}} = \mathbf{x}_i^\intercal \boldsymbol{\beta}_j, \text{ for } j = 0, \ldots, J-2, \tag{12}$$

with $p_{i,J-1} = 1 - \sum_{j=0}^{J-2} p_{ij}$. Consider the case of $j = 3$ levels and a single binary covariate x so that that $\mathbf{x}_i = \begin{pmatrix} 1 & x_i \end{pmatrix}^\intercal$. Give a 3×2 table containing the probabilities of $\mathbb{P}\left(Y = j \mid x\right)$ in terms the β_{jx}s for rows $j = 0, 1, 2$ and columns $x = 0, 1$. Hence, give interpretations of $\exp\left(\beta_{jx}\right)$ for $j = 0, 1, 2$ and $x = 0, 1$.

Is the generalized logit model suitable for ordinal data?

j	$x = 0$	$x = 1$
0	$\frac{\exp\left(\beta_{00}\right)}{1 + \exp\left(\beta_{00}\right) + \exp\left(\beta_{10}\right)}$	$\frac{\exp\left(\beta_{00} + \beta_{01}\right)}{1 + \exp\left(\beta_{00} + \beta_{01}\right) + \exp\left(\beta_{10} + \beta_{11}\right)}$
1	$\frac{\exp\left(\beta_{10}\right)}{1 + \exp\left(\beta_{00}\right) + \exp\left(\beta_{10}\right)}$	$\frac{\exp\left(\beta_{10} + \beta_{11}\right)}{1 + \exp\left(\beta_{00} + \beta_{01}\right) + \exp\left(\beta_{10} + \beta_{11}\right)}$
2	$\frac{1}{1 + \exp\left(\beta_{00}\right) + \exp\left(\beta_{10}\right)}$	$\frac{1}{1 + \exp\left(\beta_{00} + \beta_{01}\right) + \exp\left(\beta_{10} + \beta_{11}\right)}$

Table 3: Multinomial probabilities for various j and x.

Solution: See Table 3 for the table of probabilities.

Equation 12 provides a way to interpret the β_{jx}. Let p_{0j} and p_{1j} denote the probabilities when $x = 0$ and $x = 1$, respectively. In this case, we have the odds ratios:

$$\frac{p_{0j}}{p_{02}} = \exp\left(\beta_{j0}\right)$$
$$\frac{p_{1j}}{p_{12}} = \exp\left(\beta_{j0}\right) \exp\left(\beta_{j1}\right).$$

Thus, the coefficients β_{j0} are the expected log odds ratio for level j relative to level $J - 1 = 2$ when the $x = 0$. β_{j1} is the expected increase in this log odds ratio when $x = 1$. In this sense, we can consider the level $J - 1$ the default case, and $\exp(\beta_{jx})$ express how much more likely we are to observe level j.

This model isn't suitable for ordinal data, for it is agnostic to the order of the data. From the above interpretation, it's more similar to fitting $J - 1$ individual logisitic regression models. For an ordinal model, we might want behavior like the most probable level varies monotonically with some covariate. There's no way to model such behavior with the *generalized logit model* since each class has separate paramters.

5. Let

$$\pi_{ij} = \mathbb{P}\left(Y_i \leq j \mid \mathbf{x}_i\right), \tag{13}$$

for $j = 0, \ldots, J-2$ and with $\mathbf{x}_i = \begin{pmatrix} 1 & x_{1i} & x_{2i} & \cdots & x_{ki} \end{pmatrix}^\intercal$. Consider the proportional odds model

$$\log \frac{\pi_{ij}}{1 - \pi_{ij}} = \alpha_j - \mathbf{x}_i^\intercal \boldsymbol{\beta}, \tag{14}$$

for $j = 0, 1, J - 2$, and where $\boldsymbol{\beta} = \begin{pmatrix} \beta_0 & \beta_1 & \cdots & \beta_k \end{pmatrix}^\intercal$. Write down, in as simplified a form as possible, the log-likelihood $l\left(\boldsymbol{\alpha}, \boldsymbol{\beta}\right)$ where $\boldsymbol{\alpha} = \begin{pmatrix} \alpha_0 & \alpha_1 & \cdots & \alpha_{J-2} \end{pmatrix}^\intercal$, for the sample y_i, $i = 1, \ldots, n$.

Solution: