Introduction to multinomial regression

Motivating example: earthquake data

We have data from the 2015 Gorkha earthquake in Nepal. After the earthquake, a large scale survey was conducted to determine the amount of damage the earthquake caused for homes, businesses and other structures. Variables include:

- `Damage`: the amount of damage suffered by the building (none, moderate, severe)
- `age`: the age of the building (in years)
- `condition`: a de-identified variable recording the condition of the land surrounding the building

Research goal: Build a model to predict Damage

> Are any of the models we have learned so far suitable for predicting Damage? No! Damage is categorical, w/ > 2 levels

Categorical distribution = multinomial when $n=1$

The categorical distribution

Multinomial distribution : counts # of observations in each of J categories, out of n total observations

If $n=1$, then we get Categorical distribution (like how Bernoulli = Binomial($n=1$))

Damage has levels None, Moderate, Severe

For now, ignore ordering

$$Damage_i \sim Categorical(\pi_{i(None)}, \pi_{i(Moderate)}, \pi_{i(Severe)})$$

$\pi_{i(None)} = P(Damage_i = None)$, etc.

In general, $Y \sim Categorical(\pi_1, \dots, \pi_J)$ when $Y \in \{1, \dots, J\}$

w/ probabilities $\pi_j = P(Y=j)$ and $\sum_{j=1}^{J} \pi_j = 1$

$(j = 1, \dots, J)$

Write Categorical like an EDM

$Y \sim \text{Categorical}(\pi_1, \dots, \pi_J)$

Let $Y_j^* = \begin{cases} 1 & Y = j \\ 0 & Y \neq j \end{cases}$ for $j = 1, \dots, J$

$$f(y; \pi_1, \dots, \pi_J) = \begin{cases} \pi_1 & Y = 1 \\ \pi_2 & Y = 2 \\ \vdots & \\ \pi_J & Y = J \end{cases} = \prod_{j=1}^{J} \pi_j^{y_j^*}$$

$$\sum_{j=1}^{J} \pi_j = 1 \Rightarrow \pi_J = 1 - \sum_{j=1}^{J-1} \pi_j \qquad \sum_{j=1}^{J} y_j^* = 1 \Rightarrow y_J^* = 1 - \sum_{j=1}^{J-1} y_j^*$$

$$\Rightarrow f(y; \pi_1, \dots, \pi_{J-1}) = \left(\prod_{j=1}^{J-1} \pi_j^{y_j^*} \right) \left(1 - \sum_{j=1}^{J-1} \pi_j \right)^{1 - \sum_{j=1}^{J-1} y_j^*}$$

$$= \exp\left\{ \sum_{j=1}^{J-1} y_j^* \log \pi_j + \left(1 - \sum_{j=1}^{J-1} y_j^*\right) \log\left(1 - \sum_{j=1}^{J-1} \pi_j\right) \right\}$$

$$= \exp\left\{ \sum_{j=1}^{J-1} y_j^* \log\left(\frac{\pi_j}{1 - \sum_{k=1}^{J-1} \pi_k} \right) + \log\left(1 - \sum_{j=1}^{J-1} \pi_j\right) \right\}$$

$$f(y^*; \pi_1, \dots, \pi_{J-1}) = \exp\left\{ \underbrace{\sum_{j=1}^{J-1} y_j^* \log\left(\frac{\pi_j}{1-\sum_{k=1}^{J-1}\pi_k}\right)}_{y^{*T}\theta} + \underbrace{\log\left(1 - \sum_{j=1}^{J-1}\pi_j\right)}_{-\kappa(\theta)} \right\}$$

$$\kappa : \mathbb{R}^{J-1} \to \mathbb{R}$$

$\phi = 1 \in \mathbb{R}$

$a(y^*, \phi) = 1 \in \mathbb{R}$

$y^* \in \mathbb{R}^{J-1}$

$$\theta \in \mathbb{R}^{J-1} = \left(\log\left(\frac{\pi_1}{1-\sum_{k=1}^{J-1}\pi_k}\right), \log\left(\frac{\pi_2}{1-\sum_{k=1}^{J-1}\pi_k}\right), \dots, \log\left(\frac{\pi_{J-1}}{1-\sum_{k=1}^{J-1}\pi_k}\right) \right)$$

Multivariate of EDM: $\quad f(y; \theta, \phi) = a(y, \phi) \exp\left\{ \frac{y^T\theta - \kappa(\theta)}{\phi} \right\}$

If $J = 2$, then Categorical $\equiv$ Bernoulli

$X_i^* \in \mathbb{R}^{(J-1) \times ((K+1)(J-1))}$ $\beta \in \mathbb{R}^{((K+1)(J-1)) \times 1}$

Multivariate GLM

Suppose we observe data $(X_1, Y_1), \dots, (X_n, Y_n)$ $X_i \in \mathbb{R}^{K+1}$

$Y_i \sim \text{Categorical}(\pi_{i1}, \dots, \pi_{iJ})$. Let $Y_{ij}^* = \begin{cases} 1 & Y_i = j \\ 0 & Y_i \neq j \end{cases}$

$Y_i^* = (Y_{i1}^*, \dots, Y_{i,J-1}^*)^T \in \mathbb{R}^{J-1}$

$\mathbb{E}[Y_i^*] = \mu_i = (\pi_{i1}, \dots, \pi_{i,J-1})^T$

Multivariate GLM

$$g_1(\mu_i) = \beta_1^T X_i \qquad \beta_1 = (\beta_{10}, \beta_{11}, \dots, \beta_{1K})^T \in \mathbb{R}^{K+1}$$

$$g_2(\mu_i) = \beta_2^T X_i \qquad \beta_2 = (\beta_{20}, \beta_{21}, \dots, \beta_{2K})^T \in \mathbb{R}^{K+1}$$

$$\vdots$$

$$g_{J-1}(\mu_i) = \beta_{J-1}^T X_i$$

$g : \mathbb{R}^{J-1} \to \mathbb{R}^{J-1}$

$$g(\mu_i) = \begin{pmatrix} g_1(\mu_i) \\ \vdots \\ g_{J-1}(\mu_i) \end{pmatrix} = \begin{bmatrix} \beta_1^T X_i \\ \vdots \\ \beta_{J-1}^T X_i \end{bmatrix} = \underbrace{\begin{bmatrix} X_i^T & & & \\ & X_i^T & & \\ & & \ddots & \\ & & & X_i^T \end{bmatrix}}_{X_i^*} \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{J-1} \end{bmatrix}}_{\beta} = X_i^* \beta$$

Multinomial regression model

For multinomial regression model, the canonical link would be

$$g(\mu_i) = \theta_i = \begin{pmatrix} \log\left(\frac{\hat{\pi}_1}{1-\sum_{j=1}^{J-1}\hat{\pi}_j}\right) \\ \vdots \\ \log\left(\frac{\hat{\pi}_{J-1}}{1-\sum_{j=1}^{J-1}\hat{\pi}_j}\right) \end{pmatrix}$$

$$\hat{\pi}_J = 1 - \sum_{j=1}^{J-1} \hat{\pi}_j$$

written another way:

$$\left.\begin{aligned} \log\left(\frac{\hat{\pi}_1}{\hat{\pi}_J}\right) &= \beta_1^T X_i \\ \log\left(\frac{\hat{\pi}_2}{\hat{\pi}_J}\right) &= \beta_2^T X_i \\ &\vdots \\ \log\left(\frac{\hat{\pi}_{J-1}}{\hat{\pi}_J}\right) &= \beta_{J-1}^T X_i \end{aligned}\right\} \text{baseline-category logits}$$

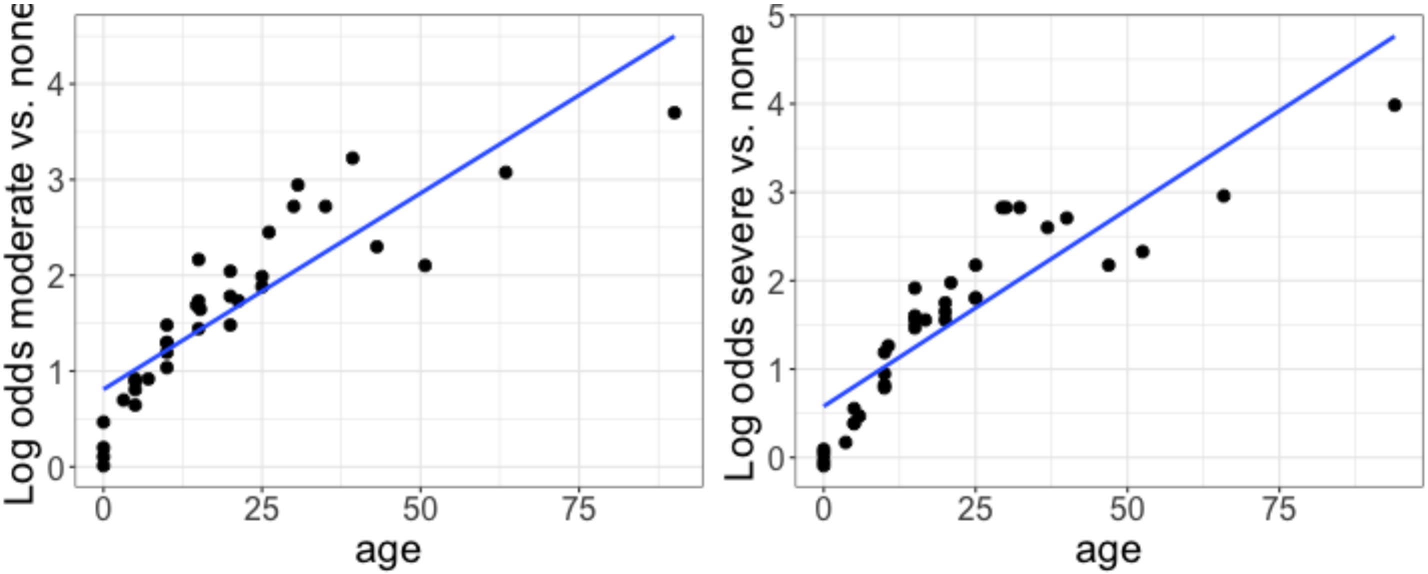
Multinomial regression model

$$\begin{cases} \text{Damage}_i \sim \text{Categorical}(\pi_{i(\text{None})}, \pi_{i(\text{moderate})}, \pi_{i(\text{severe})}) \\ \log\left(\dfrac{\pi_{i(\text{Moderate})}}{\pi_{i(\text{None})}}\right) = \beta^T_{(\text{moderate})} X_i \\ \log\left(\dfrac{\pi_{i(\text{severe})}}{\pi_{i(\text{None})}}\right) = \beta^T_{(\text{Severe})} X_i \end{cases}$$

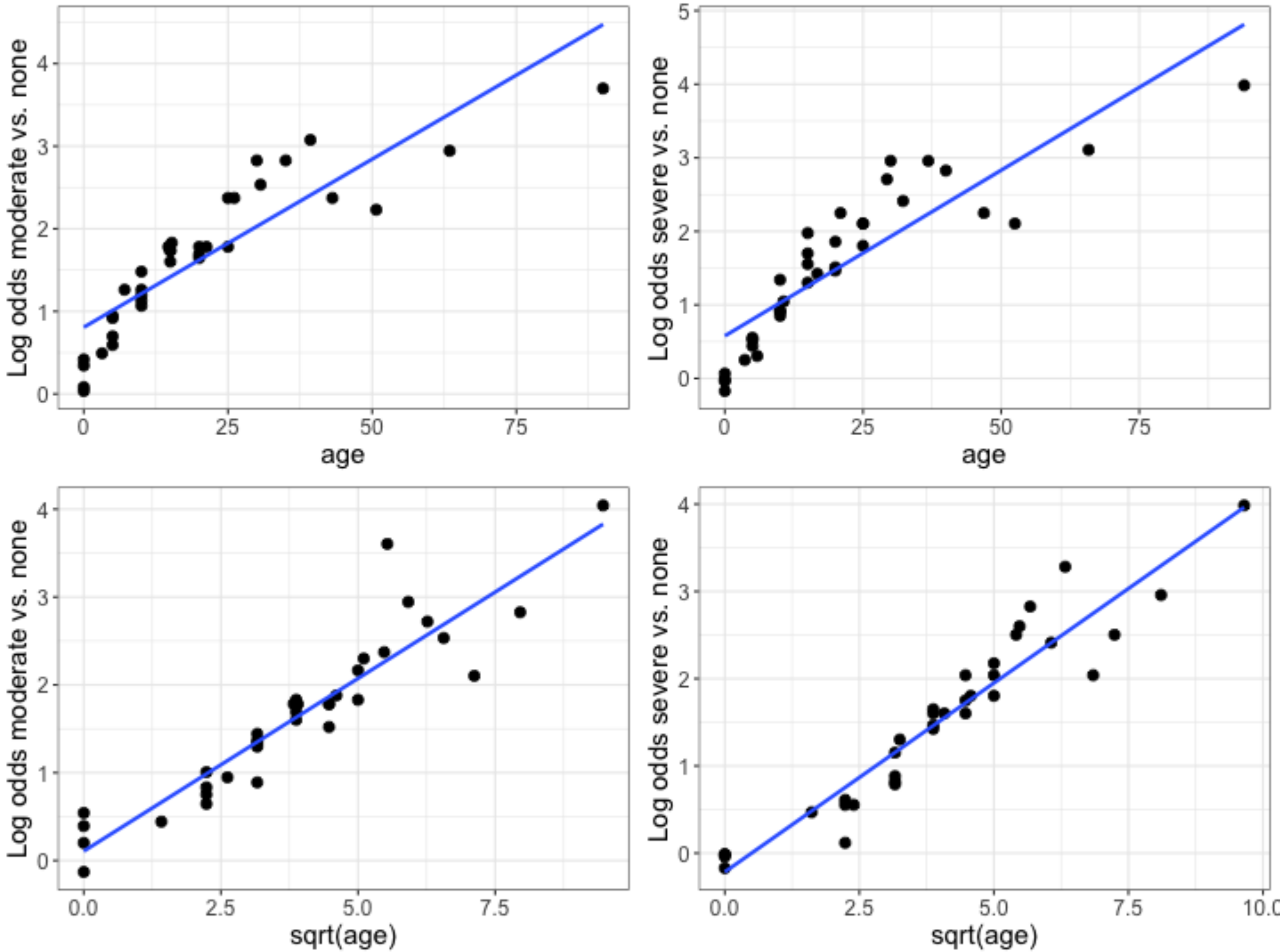
Exploratory data analysis

> We want to model damage using age and land surface condition. What kind of EDA could I do?

Empirical logit plots



Trying a transformation



Fitting the model in R

```
library(nnet)
m1 <- multinom(Damage ~ sqrt(age) +
                  condition,
                data = earthquake)
```

```
summary(m1)
```

```
...
## Coefficients:
##            (Intercept) sqrt(age)  conditiono conditiont
## moderate   0.6581163 0.3747641 -0.45376940 -0.5803708
## severe     0.1881145 0.4251732  0.04706934 -0.4623774
##
## Std. Errors:
##            (Intercept)  sqrt(age) conditiono conditiont
## moderate   0.1208913 0.01684468  0.2305975  0.1155475
## severe     0.1243799 0.01725782  0.2292533  0.1180182
...
```

Class activity

https://sta712-f22.github.io/class_activities/ca_lecture_34.html