Multinomial regression

Recap: multinomial regression model

data $(x_1, y_1), \ldots, (x_n, y_n)$ $\qquad x_i \in \mathbb{R}^{(k+1)}$

$Y_i \sim \text{Categorical}(\pi_{i1}, \ldots, \pi_{iJ})$

$\mu_i = (\pi_{i1}, \ldots, \pi_{i,J-1})^T \in \mathbb{R}^{J-1}$

$$g(\mu_i) = \begin{pmatrix} \log\left(\dfrac{\pi_{i1}}{1 - \sum_{j=1}^{J-1} \pi_{ij}}\right) \\ \vdots \\ \log\left(\dfrac{\pi_{i,J-1}}{1 - \sum_{j=1}^{J-1} \pi_{ij}}\right) \end{pmatrix} = \begin{pmatrix} \beta_1^T x_i \\ \vdots \\ \beta_{J-1}^T x_i \end{pmatrix} = x_i^* \beta$$

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{J-1} \end{pmatrix} \in \mathbb{R}^{(J-1)(k+1)}$$

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

$$\text{Damage}_i \sim \text{Categorical}(\pi_{i(\text{None})}, \pi_{i(\text{Moderate})}, \pi_{i(\text{Severe})})$$

$$\log\left(\frac{\pi_{i(\text{Moderate})}}{\pi_{i(\text{None})}}\right) = \beta_{(\text{Moderate})}^T X_i$$

$$\log\left(\frac{\pi_{i(\text{Severe})}}{\pi_{i(\text{None})}}\right) = \beta_{(\text{Severe})}^T X_i$$

Exploratory data analysis

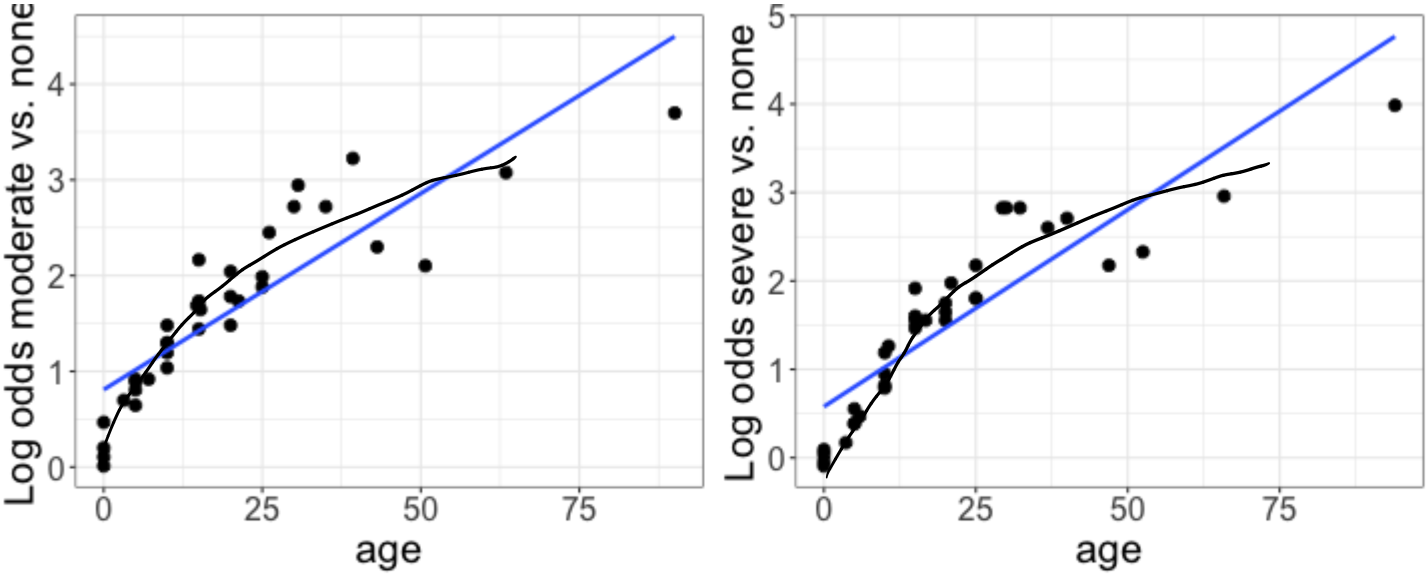
> We want to model damage using age and land surface condition. What kind of EDA could I do?

Empirical logit plots!

Compare Moderate vs. None

Severe vs. None

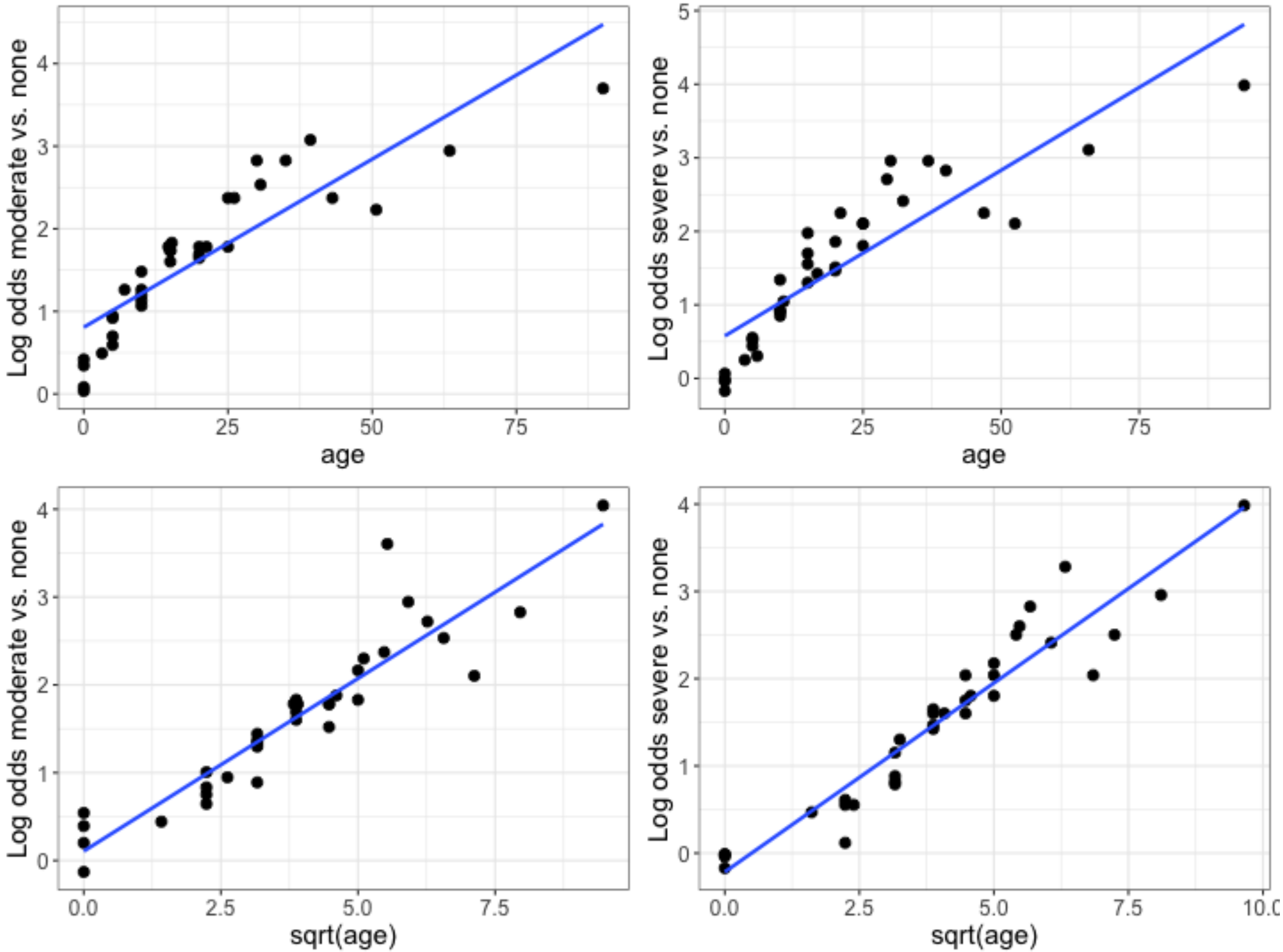
Empirical logit plots



- Similar relationship w/ age in each plot
- relationship looks slightly nonlinear

$\Rightarrow$ transformation?

Trying a transformation



Fitting the model in R

package w/ multinom function

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

```r
summary(m1)
```

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

moderate vs. none

severe vs. none

A one-unit increase in $\sqrt{Age}$ is associated with an increase in the odds of moderate vs. no damage by a factor of $\exp\{0.375\} = 1.45$, holding surface condition fixed

Class activity

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

Class activity

1) $\exp\{0.658 + 0.375\sqrt{25} - 0.454\}$

$\exp\{2.08\} \approx 8 = \dfrac{\hat{\pi}_{i(\text{Moderate})}}{\hat{\pi}_{i(\text{None})}}$

2) $\hat{\pi}_{i(\text{Moderate})} = \dfrac{\hat{\pi}_{i(\text{Mod})} / \hat{\pi}_{i(\text{None})}}{1 + \dfrac{\hat{\pi}_{i(\text{Mod})}}{\hat{\pi}_{i(\text{None})}} + \dfrac{\hat{\pi}_{i(\text{sev.})}}{\hat{\pi}_{i(\text{None})}}}$

$= \dfrac{\hat{\pi}_{i(\text{mod})}}{\hat{\pi}_{i(\text{None})} + \hat{\pi}_{i(\text{mod})} + \hat{\pi}_{i(\text{sev.})}}$

In general:

$\hat{\pi}_{ij} = \dfrac{\text{odds}\,(j \text{ vs. } J)}{1 + \sum_{k=1}^{J-1} \text{odds}\,(k \text{ vs. } J)}$

$= \dfrac{\exp\{2.08\}}{1 + \exp\{2.08\} + \exp\{2.36\}} \approx 0.4$

Fisher scoring for multinomial regression

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

$\Rightarrow \log f(y;\theta,\phi) = \log a(y,\phi) + \frac{1}{\phi}\left(y^T\theta - \kappa(\theta)\right)$

$$\boxed{Y^*_{ij} = \begin{cases} 1 & Y_i = j \\ 0 & Y_i \neq j \end{cases}}$$

multivariate GLM: $\quad g(\mu_i) = X_i^* \beta$

$$\ell(\beta) = \sum_{i=1}^{n} \log a(Y_i^*, \phi) + \frac{1}{\phi}\sum_{i=1}^{n}\left(Y_i^{*T}\theta_i - \kappa(\theta_i)\right)$$

$$u(\beta) = \frac{\partial \ell}{\partial \beta} = \frac{1}{\phi}\sum_{i=1}^{n}\frac{\partial}{\partial\beta}\left(\underbrace{Y_i^{*T}\theta_i}_{Y_i^{*T}X_i^*\beta} - \kappa(\theta_i)\right)$$

canonical link: $\quad \theta_i = g(\mu_i) = X_i^*\beta \Rightarrow Y_i^{*T}X_i^*\beta$

$$\frac{\partial}{\partial\beta} Y_i^{*T}X_i^*\beta = \frac{\partial}{\partial\beta}\beta^T X_i^{*T} Y_i^* = X_i^{*T} Y_i^*$$

$$\frac{\partial}{\partial\beta}\kappa(\theta_i) \overset{\text{(chain rule)}}{=} \frac{\partial X_i^*\beta}{\partial\beta}\,\frac{\partial\kappa(\theta_i)}{\partial\theta_i} = X_i^{*T}\mu_i$$

$$\Rightarrow U(\beta) = \frac{1}{\phi} \sum_{i=1}^{n} X_i^{*T} (Y_i^* - \mu_i)$$

$$X_i^* = \begin{bmatrix} X_i^T & & & \\ & X_i^T & & \\ & & \ddots & \\ & & & X_i^T \end{bmatrix}$$