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?

The categorical distribution

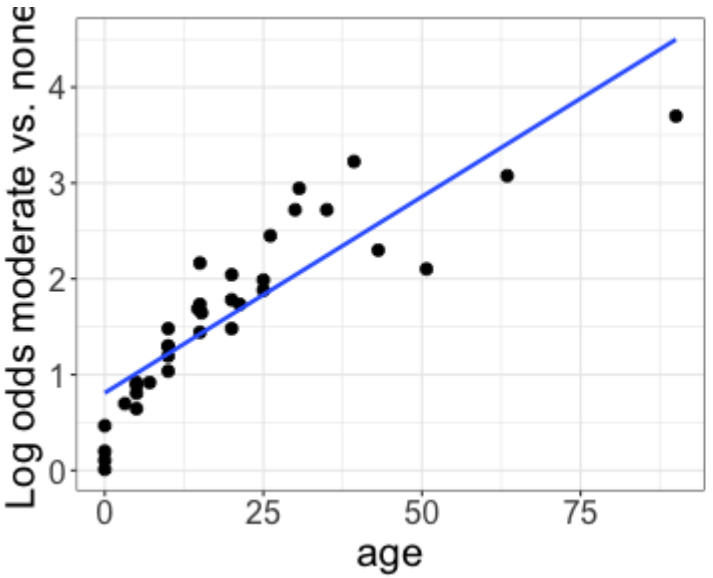
Multivariate GLM

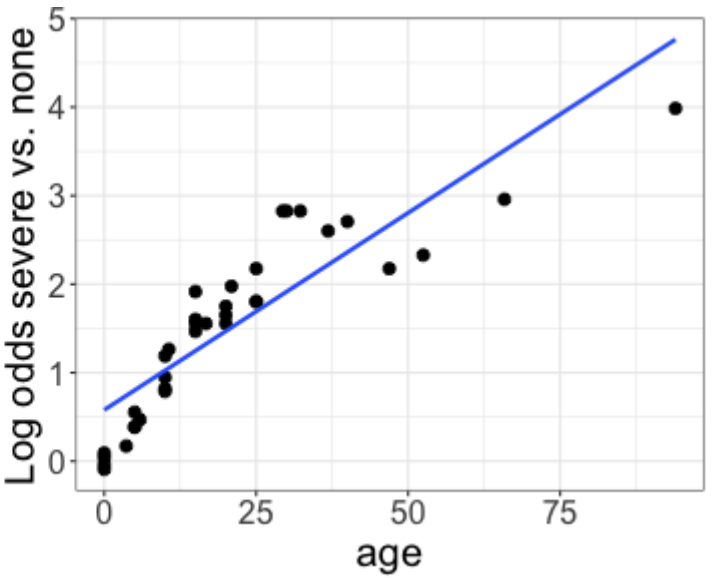
Multinomial regression model

Exploratory data analysis

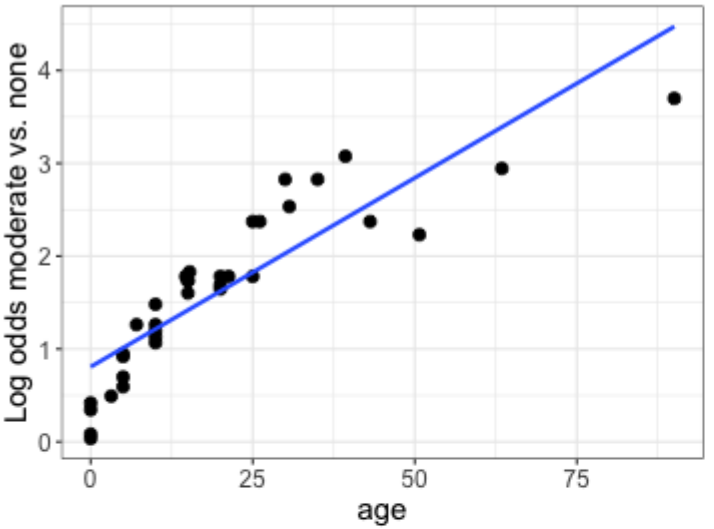
> We want to model damage using age and land surface condition. What kind of EDA could I do?

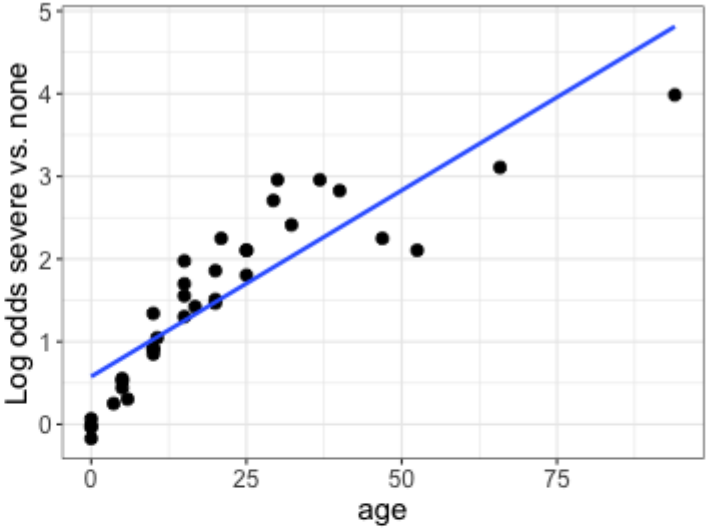
Empirical logit plots

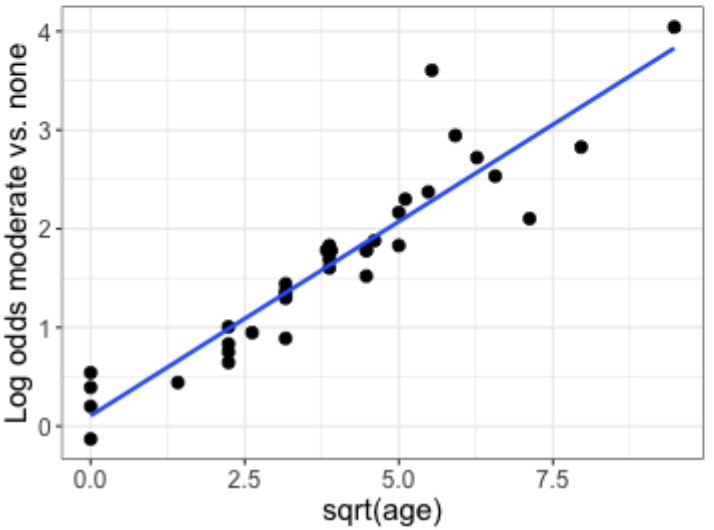


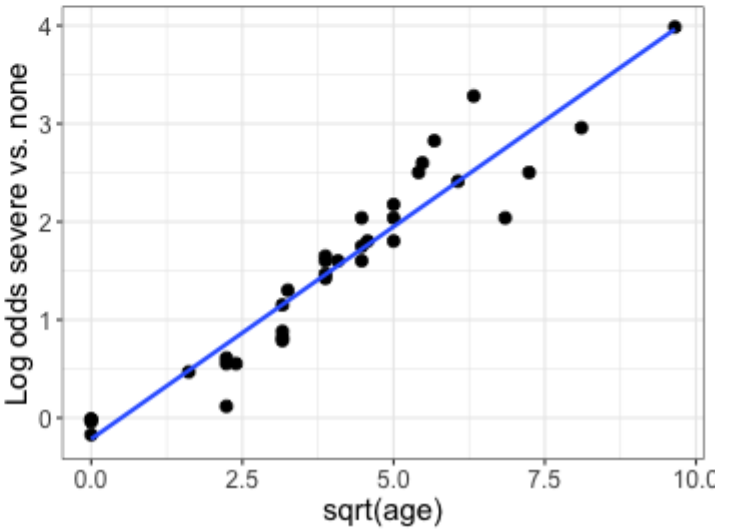


Trying a transformation









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