Lab: Logistic Regression

1. Field goal success probability

We have a dataset consisting of field goals,

- row = a field goal
- i = index of i^{th} field goal attempt
- y_i = 1 if field goal made else 0
- ydl_i = yardline (yards from opp. endzone) of i^{th} kick
- kq_i = kicker quality of the i^{th} punter (I made this variable)
- $Kicker_i$ = the name of the kicker

- Model field goal success probability
- use out-of-sample predictive performance to select a model if you don't know which of several models to use
- visualize the model(s)

2. Bradley Terry NCAA Mens Basketball Power Scores

We have a dataset consisting of the results of each game in a college mens basketball season,

- row = a game
- i = index of i^{th} game
- h_i = index of home team
- a_i = index of away team
- y_i = 1 if home team wins else 0

The Bradley Terry model supposes each team j has a latent power rating or strength β_j and the probability that team j beats team k at home is

$$p_{jk} = \text{Logistic}(\beta_0 + \beta_j - \beta_k)$$

and on the road is $1 - p_{kj} = 1 - \text{Logistic}(\beta_0 + \beta_k - \beta_j)$.

So, β_0 is a home field advantage parameter.

The Bradley Terry model is just logistic regression with a strength parameter for each team.

Fit the model, visualize the coefficients, and interpret the home field advantage parameter.

Note: **ELO Power Scores**

ELO is an "online" or rolling version of Bradley Terry Logistic Regression Power scores, updated after every match.

Think chess or tennis.
Player i's ELO is β_i.

Model $P_{ij} = \mathbb{P}(i \text{ beats } j) = \dfrac{1}{1 + e^{\beta_j - \beta_i}}$

If i beats j, update:

$$\beta_i \leftarrow \beta_i + K(1 - P_{ij})$$
$$\beta_j \leftarrow \beta_j - K(1 - P_{ij})$$