Lab: Multivariable Linear Regression

1. Expected outcome of a punt

We have a dataset consisting of punts,

$$\begin{cases} \text{row} = \text{a punt} \\ i = \text{index of } i\text{th punt} \\ y_i = \text{outcome (next yardline, from opponent's perspective) of the punt} \\ ydl_i = \text{yardline (yards from opp. endzone) of } i\text{th kick} \\ pq_i = \text{Punter quality of the } i\text{th punter} \quad (\text{I made this variable}) \\ punter_i = \text{name of punter} \end{cases}$$

- Model the outcome (next yardline) after a punt by

$$\begin{cases} y_i = \beta_0 + \beta_1 \cdot ydl_i + \varepsilon_i \\ y_i = \beta_0 + \text{spline}(ydl_i / \alpha) + \varepsilon_i \\ y_i = \beta_0 + \text{spline}(ydl_i / \alpha) + \beta_i \cdot pq_i + \varepsilon_i \end{cases}$$

and try other variations $\left(e.g. \text{ quadratic, cubic in yardline}\right)$.

- Compare the out-of-sample predictive performance of these 3 models

- Visualize the best model