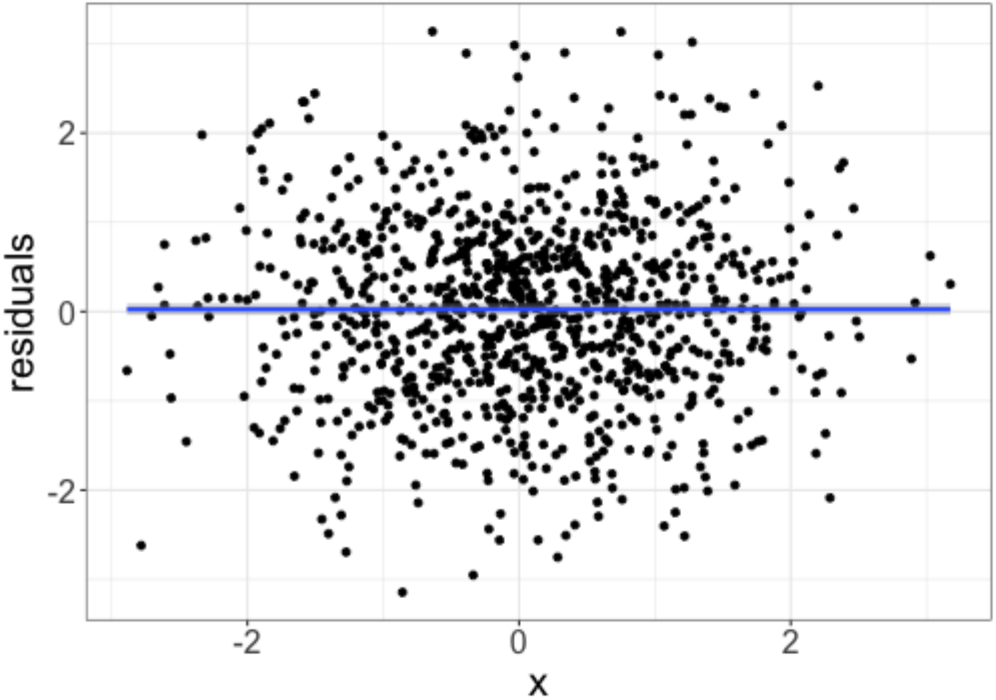
Logistic regression assumptions and diagnostics

- HW 1 due Tuesday

- Challenge 2 released on course website
  - Logistic regression "in the wild"

Last time: quantile residuals to assess shape



Warm up: Class activity, Part I

https://sta712-f22.github.io/class_activities/ca_lecture_6.html

- Generate data for which the logistic regression shape assumption doesn't hold
- See whether the violation shows up on a quantile residual plot

More logistic regression diagnostics

- Are there any outliers that could affect the fitted model?
- Are there issues with multicollinearity?

correlation
variance inflation
factors

leverage &
Cook's distance

Leverage and Cook's Distance in linear regression

Linear regression: $Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$ $X = \begin{pmatrix} 1 & X_{11} & \cdots & X_{1k} \\ 1 & X_{21} & & \\ \vdots & \vdots & & \\ 1 & X_{n1} & \cdots & X_{nk} \end{pmatrix}$

$Y = X\beta + \varepsilon$ $\varepsilon \sim N(0, \sigma^2 I)$ $\beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix}$

$\Rightarrow \hat{Y} = \underbrace{X(X^T X)^{-1} X^T}_{\text{"hat matrix" } H} Y$

leverage of observation i (i-th diagonal element of H)

$\text{var}(Y - \hat{Y}) = \sigma^2 (I - H)$ $\Rightarrow \text{Var}(Y_i - \hat{Y}_i) = \sigma^2 (1 - h_i)$

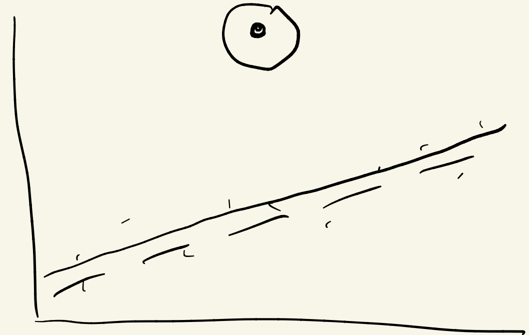
$h_i \leq 1$ $\uparrow h_i \Rightarrow \text{var}(Y_i - \hat{Y}_i) \downarrow$

SLR: $h_i = \frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum_j (X_j - \bar{X})^2}$

outlier, high leverage

$\downarrow$

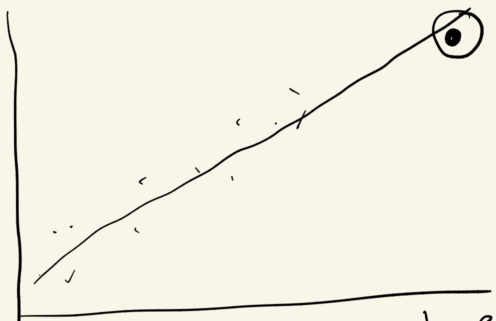
influential



outlier, low leverage

$\downarrow$

not influential



not outlier, high leverage

$\Rightarrow$ not influential

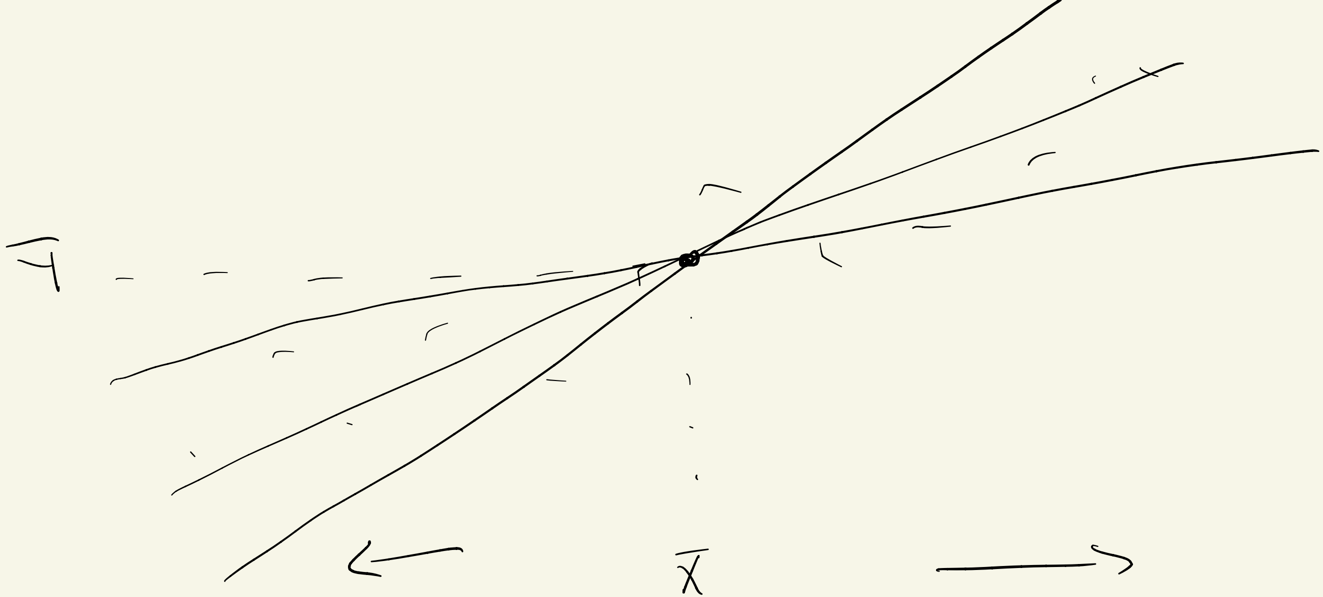
Cook's Distance

$$D_i = \frac{(Y_i - \hat{Y}_i)^2}{\hat{\sigma}^2} \cdot \frac{h_i}{(1-h_i)^2}$$

Concerned when $D_i >$ threshold

e.g. 0.5

or 1


Y
X

Fisher scoring as Iteratively Reweighted Least Squares

Fisher scoring : $\beta^{(r+1)} = \beta^{(r)} + \mathcal{I}^{-1}(\beta^{(r)})\, u(\beta^{(r)})$

HW 1, #3 $\Rightarrow$ $u(\beta) = X^T(Y - p)$

$$p = \left[\underbrace{\frac{e^{\beta^T x_1}}{1+e^{\beta^T x_1}}}_{p_1}, \underbrace{\frac{e^{\beta^T x_2}}{1+e^{\beta^T x_2}}}_{p_2}, \ldots, \frac{e^{\beta^T x_n}}{1+e^{\beta^T x_n}}\right]^T$$

$\mathcal{I}(\beta) = X^T W X$ $\quad$ $W = \mathrm{diag}(p_1(1-p_1), p_2(1-p_2), \ldots, p_n(1-p_n))$

$$= \beta^{(r+1)} = \beta^{(r)} + (X^T W^{(r)} X)^{-1} X^T (Y - p^{(r)})$$

$$\beta^{(r+1)} = \beta^{(r)} + \left(X^T W^{(r)} X\right)^{-1} X^T (Y - p^{(r)})$$

$$= \underbrace{\left(X^T W^{(r)} X\right)^{-1} X^T W^{(r)} X}_{I} \beta^{(r)} + \left(X^T W^{(r)} X\right)^{-1} X^T (Y - p^{(r)})$$

$$= \left(X^T W^{(r)} X\right)^{-1} X^T W^{(r)} \underbrace{\left(X\beta^{(r)} + \left(W^{(r)}\right)^{-1} (Y - p^{(r)}) \right)}_{Z^{(r)}}$$

$\leftarrow$ working responses (at iteration r)

$$\Rightarrow \beta^{(r+1)} = \left(X^T W^{(r)} X\right)^{-1} X^T W^{(r)} Z^{(r)}$$

$\Rightarrow$ Fisher scoring is really weighted least squares with weights $W^{(r)}$ and response $Z^{(r)}$

Call this iteratively reweighted least squares (IRLS)

$Y = X\beta + \varepsilon$

what if ε doesn't have constant variance?

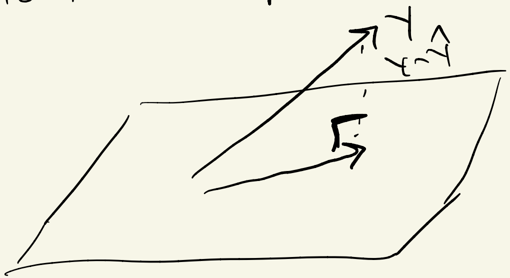
Try to minimize $\sum_{i=1}^{n} w_i (Y_i - \hat{Y}_i)^2$

usually: $w_i = \frac{1}{\mathrm{Var}(Y_i)} = \frac{1}{\mathrm{Var}(\varepsilon_i)}$

(inverse variance weighting)

$W = \mathrm{diag}(W_1, W_2, \ldots, W_n)$

want to find $\hat{Y}$ in colspace of WX that is "close to" Y



$\Rightarrow X^T W (Y - \hat{Y}) = 0$

$\Rightarrow X^T W Y = X^T W \hat{Y}$ $\qquad \hat{Y} = X\hat{\beta}$

$X^T W Y = X^T W X \hat{\beta}$

$\Rightarrow X^T W Y = X^T W X \hat{\beta}$

$\Rightarrow \hat{\beta} = (X^T W X)^{-1} X^T W Y$

$$\beta^{(r+1)} = (X^T W^{(r)} X)^{-1} X^T W^{(r)} Z^{(r)}$$

$$\hat{Z}^{(r+1)} = X \beta^{(r+1)}$$

$$= \underbrace{X (X^T W^{(r)} X)^{-1} X^T W^{(r)}}_{H} Z^{(r)}$$

Class Activity, Part II

https://sta712-f22.github.io/class_activities/ca_lecture_6.html

> Exploring leverage and Cook's distance!