Setup $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$

given obs $\underline{y}$ we have

fitted values: $\hat{\underline{y}} = X\hat{\underline{\beta}}$

estimated residuals: $\hat{\underline{\varepsilon}} = \underline{y} - \hat{\underline{y}} = \underline{y} - X\hat{\underline{\beta}}$

where $\hat{\underline{\beta}} = (X^T X)^{-1} X^T \underline{y}$

DEF The hat matrix is

$$H = X(X^T X)^{-1} X^T$$

note this: idempotent!

$$H^2 = HH = X(X^TX)^{-1}X^T \, X(X^TX)^{-1}X^T = X(X^TX)^{-1}X^T = H$$

The fitted values are

$$\hat{Y} = X\hat{\beta} = X(X^TX)^{-1}X^TY = HY.$$

So H projects data onto fitted values.

Estimated resid are

$$\hat{\varepsilon} = Y - \hat{Y} = Y - HY = (I - H)Y.$$

Thus, $\boxed{\text{Var } \hat{\varepsilon} = \sigma^2 (I - H)}$ [proof: HW]

So, the SE for $\hat{\varepsilon}_i$ is

$$SE(\hat{\varepsilon}_i) = \hat{\sigma}\sqrt{1 - H_{ii}}$$

H_{ii} → ith element of $\text{diag}(H)$

Under A1, the <u>studentized residuals</u> are

$$\hat{\varepsilon}_i^* = \frac{\hat{\varepsilon}_i - 0}{SE(\hat{\varepsilon}_i)} \sim N(0,1)$$

which gives us a quantitative way to assess outliers, i.e., if $\hat{\varepsilon}_i^*$ is outside of $[-2, 2]$ then it is potentially an outlier.

[in R "standardized residuals" = "studentized residuals"]

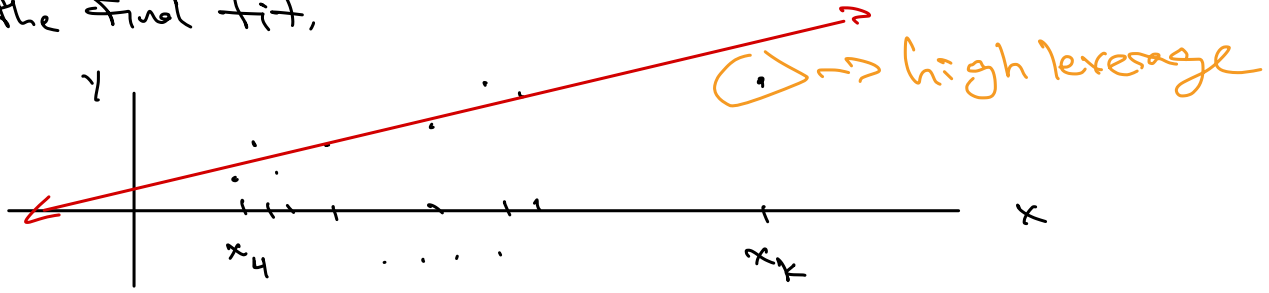
Note $\hat{y} = H y$

$\Rightarrow \hat{y}_i = H_{i1} y_1 + \cdots + H_{ii} y_i + \cdots + H_{in} y_n$

so H_{ii} is the influence of y_i on its own fitted value $\hat{y}_i$

DEF Define H_{ii} to be the leverage of (y_i, x_i)

Note Leverage is the potential for y_i having an influence on the final fit.



data points with unusual x values have high leverage

Remark: $H = X(X^TX)^{-1}X^T$ only depends on X, not the actual value of y_i, so it only measures potential influence.

To measure actual influence, Cook's Distance (Cook's D) is

$$D_i = \frac{H_{ii}}{p(1-H_{ii})}(\hat{\varepsilon}_i^*)^2$$

leverages are less than 1

Note:

High leverage $\Rightarrow$ potential for influencing fit

High Cook's D $\Rightarrow$ actual influence over fit

[R]

Assessing quality of model fit

How do we compare competing models, i.e., assess goodness-of-fit?

Recall

$$RSS = (y - X\hat{\beta})^T (y - X\hat{\beta}) = \sum_i (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_p x_{ip}))^2$$

and

$$r^2 = 1 - \frac{RSS}{SYY} \quad \text{where } SYY = \sum (y_i - \bar{y})^2$$

Note $r^2 \in [0,1]$, larger values $\Rightarrow$ better fit

RSS, smaller values $\Rightarrow$ better fit.

But: Both will always improve with the introduction of more features.

If model has d features, define

Mallow's C_p : $C_p = \frac{1}{n}(RSS + 2d\hat{\sigma}^2)$

Akaike information criterion (AIC) : $AIC = \frac{1}{n\hat{\sigma}^2}(RSS + 2d\hat{\sigma}^2)$

Bayesian information criterion (BIC) : $BIC = \frac{1}{n}(RSS + \log(n)\, d\hat{\sigma}^2)$

We prefer models with smaller C_p / AIC / BIC.

BIC tends to choose simpler models than AIC.