Chapter 8

Regression Diagnostic - Residuals

8.1. Regression Diagnostics

$$Y = X\beta + e, \quad e \sim N(0, \sigma^2)$$

- Regression diagnostic
  - Check if the assumptions (mean/var/error) are consistent with the observed data.
- Study the residuals
  - If the model works well, then the residuals matches the assumption
    - Mean zero, constant variance
    - i.e. residual plot looks like a null plot
      - what X cannot explain is random noise (no valid info)

8.1. Residuals

- Relationship between residuals and error

$$
\begin{aligned}
\hat{e} = Y - \hat{Y} &= \left((Y_1 - \hat{Y}_1)\ \ (Y_1 - \hat{Y}_1) \ldots (Y_n - \hat{Y}_n)\right)^T \\
&= Y - X\hat{\beta} \\
&= Y - X(X'X)^{-1}X'Y \\
&= (1 - X(X'X)^{-1}X')Y \\
&= (1 - H)Y, \qquad \text{where } H = X(X'X)^{-1}X' \text{ is the Hat matrix} \\
&= (1 - H)e, \qquad \text{since } (1 - H)X\beta = X\beta - X(X'X)^{-1}X'X\beta = 0
\end{aligned}
$$

- To study the property of residual, we need to study the property of H.

8.1. Residuals

- The Hat matrix $H = X(X'X)^{-1}X'$
  - H produces $\hat{Y}$ from Y

$$\hat{Y} = X\hat{\beta} = X(X'X)^{-1}X'Y = HY$$

- Interpretation of projection
  - H projects Y onto the space of X
  - i.e. HY can be spanned (explained) by columns of X, using the weight $\hat{\beta}$:

$$\hat{Y} = HY = X\hat{\beta} = \hat{\beta}_0 \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} + \hat{\beta}_1 \begin{pmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1n} \end{pmatrix} + \ldots + \hat{\beta}_p \begin{pmatrix} x_{p1} \\ x_{p2} \\ \vdots \\ x_{pn} \end{pmatrix}$$

8.1. Residuals

- Properties of $H = X(X'X)^{-1}X'$
  - H is symmetric.
  - $HH=?$
  - $HX=?$
  - $X'H=?$
  - $(I-H)X=?$
  - $H(I-H)=?$

8.1. Residuals

- Properties of H
  - H is symmetric. $[X(X'X)^{-1}X']$
  - HH=H $[X(X'X)^{-1}X'X(X'X)^{-1}X' = X(X'X)^{-1}X' = H]$
    - HY is already projected, so HHY=HY, so HH=H.
  - HX=X $[X(X'X)^{-1}X'X = X]$
    - Project X on X shouldn't change
  - X'H=X' $[X'X(X'X)^{-1}X' = X']$
  - (I-H)X=0 $[(I-H)X=X-HX=X-X=0]$
  - H(I-H)=0 $[H-HH=H-H=0]$

8.1. Residuals

- More properties of H
  - tr(H)= $\sum_{i=1}^{n} h_{ii}$=p+1

$$H = \begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & h_{22} & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ h_{n1} & \cdots & \cdots & h_{nn} \end{pmatrix} \qquad X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{p1} \\ 1 & x_{12} & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{pn} \end{pmatrix}$$

$$tr(H) = tr(X(X'X)^{-1}X') = tr((X'X)^{-1}X'X) = tr(I_{p+1}) = p+1$$

  - $\sum_{i=1}^{n} h_{ji} = 1, \quad \text{all j}$

$$\begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & h_{22} & \vdots & h_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n1} & \cdots & \cdots & h_{nn} \end{pmatrix}$$

Sum =1

    - Proof:
      - Check the first column of $HX = X$

Sum =1

$$\begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & h_{22} & \vdots & h_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n1} & \cdots & \cdots & h_{nn} \end{pmatrix}$$

  - $\sum_{i=1}^{n} h_{ij} = 1, \quad \text{all j}$
    - Proof:
      - Check the first row of $X'H = X'$, or use symmetry of H

8.1. Residuals

- Summary of properties of H
  1. H is symmetric.
  2. HH=H
  3. HX=X
  4. X'H=X'
  5. (I-H)X=0
  6. H(I-H)=0
  7. tr(H)=p+1
  8. $\sum_{i=1}^{n} h_{ji} = \sum_{i=1}^{n} h_{ij} = 1, \quad \text{all j}$

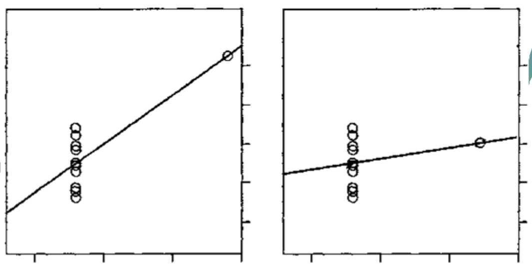
8.1. Residuals

- Remark:
  - Another use of H → Diagnostic check for leverage

- The Leverage: h_{ii} = The (i,i)-th entry of H
  - meaning: $\hat{Y} = HY$

$$\Rightarrow \hat{Y}_i = \sum_{k=1}^{n} h_{ik} Y_k = h_{ii} Y_i + \sum_{k \neq i}^{n} h_{ik} Y_k$$

  - As h_{ii} approaches 1, $\hat{Y}_i$ get closer to Y_i
    - h_{ii} is pulling $\hat{Y}_i$ towards Y_i, giving the name Leverage
    - Be careful of leverage point if h_{ii}~1



8.1. Residuals

- With properties of H, we can study property of $\hat{e}$

$$\hat{e} = Y - X\hat{\beta} = Y - X(X'X)^{-1}X'Y = (1-H)Y = (1-H)e$$

- Probabilistic properties of $\hat{e}$
  - Expectation

$$E(\hat{e}) = (I-H)E(e) = 0$$

  - Variance

$$\begin{aligned} Var(\hat{e}) &= (I-H)Var(e)(I-H)' \\ &= \sigma^2(I-H)I(I-H) \\ &= \sigma^2(I-H) \end{aligned}$$

  - The higher the leverage, the smaller the variance of $\hat{e}_i$

8.1. Residuals

- What are the difference between them?
  - $$\hat{e} \cdot 1 = 1^T \cdot \hat{e} = \sum_{i=1}^{n} \hat{e}_i = 0$$
  - $$E(e) = 0, \qquad Var(e) = \sigma^2 I$$
  - $$E(\hat{e}) = 0, \qquad Var(\hat{e}) = \sigma^2 (I - H)$$

8.1. Residuals

- What are the difference between them?
  - $$\hat{e}^T \cdot 1 = 1^T \cdot \hat{e} = \sum_{i=1}^{n} \hat{e}_i = 0$$
    - About residuals but not noise
    - Not a probabilistic property, it always holds exactly.
    - A 'by product' of finding the best fit line using Least sq
  - $$E(e) = 0, \qquad Var(e) = \sigma^2 I$$
    - A probabilistic property for noise
    - Noise have expected value 0, and are i.i.d. distributed.
  - $$E(\hat{e}) = 0, \qquad Var(\hat{e}) = \sigma^2 (I - H)$$
    - A probabilistic property for residuals
    - Residuals have expected value 0, but are
      - Dependent, and Non-identically distributed.

Powerful calculations using H

- $Cov(\hat{e},\hat{Y}) = Cov((I-H)Y, HY) = E\left(\left[(I-H)(Y-E(Y)\right]\left[H(Y-E(Y))\right]'\right)$
  $= (I-H)E\left(\left[Y-E(Y)\right]\left[Y-E(Y)\right]'\right)H' = (I-H)\sigma^2 I \cdot H' = \sigma^2 (I-H)H = 0$
- $Cov(\hat{e},Y) = Cov((I-H)Y, Y)$
  $= (I-H)\sigma^2 I = \sigma^2 (I-H) \neq 0$
- $Cov(e,Y) = ?$
- $Cov(e,\hat{Y}) = ?$

Powerful calculations using H

- $$\sum(Y_i-\hat{Y}_i)^2=(Y-\hat{Y})'(Y-\hat{Y})=[(I-H)Y]'[(I-H)Y]=Y'(I-H)Y$$
- $$\begin{pmatrix}\bar{Y}\\ \vdots\\ \bar{Y}\end{pmatrix}=\frac{1}{n}\begin{pmatrix}1 & 1 & \cdots & 1\\ 1 & 1 & \cdots & 1\\ \vdots & \vdots & \ddots & \vdots\\ 1 & \cdots & \cdots & 1\end{pmatrix}\begin{pmatrix}Y_1\\ Y_2\\ \vdots\\ Y_n\end{pmatrix}\overset{def}{=}\frac{1}{n}JY$$
- $JJ=?\quad HJ=?\quad JH=?$
- $$\sum(Y_i-\bar{Y})^2=\left(Y-\frac{1}{n}JY\right)'\left(Y-\frac{1}{n}JY\right)=Y'\left(I-\frac{1}{n}J\right)'\left(I-\frac{1}{n}J\right)Y=Y'\left(I-\frac{1}{n}J\right)Y$$
- $$\sum(\hat{Y}_i-\bar{Y})^2=\left(HY-\frac{1}{n}JY\right)'\left(HY-\frac{1}{n}JY\right)=Y'\left(H-\frac{1}{n}J\right)'\left(H-\frac{1}{n}J\right)Y=Y'\left(H-\frac{1}{n}J\right)Y$$

Powerful calculations using H

- Which is correct?

1) $\sum(Y_i - \hat{Y}_i)^2 = \sum(Y_i - \overline{Y})^2 + \sum(\hat{Y}_i - \overline{Y})^2$

2) $\sum(Y_i - \overline{Y})^2 = \sum(Y_i - \hat{Y}_i)^2 + \sum(\hat{Y}_i - \overline{Y})^2$

3) $\sum(\hat{Y}_i - \overline{Y})^2 = \sum(Y_i - \overline{Y})^2 + \sum(Y_i - \hat{Y}_i)^2$

Powerful calculations using H

- TSS=RSS+SSreg

$$\sum (Y_i - \bar{Y})^2$$

$$= Y'\left(I - \frac{1}{n}J\right)Y$$

$$= Y'\left(I - H + H - \frac{1}{n}J\right)Y$$

$$= Y'(I - H)Y + Y'\left(H - \frac{1}{n}J\right)Y$$

$$= \sum (Y_i - \hat{Y}_i)^2 + \sum (\hat{Y}_i - \bar{Y})^2$$

8.1 More powerful calculations…

$$\mathrm{E}\left(\sum (Y_i - \hat{Y}_i)^2\right) = \mathrm{E}\left(Y'(I-H)Y\right) = \mathrm{E}\left(tr\left[Y'(I-H)Y\right]\right)$$

$$= \mathrm{E}\left(tr\left[(I-H)YY'\right]\right) = tr\left[(I-H)E(YY')\right]$$

$$= tr\left[(I-H)\left(\sigma^2 I + X\beta\beta' X'\right)\right] \qquad \mathrm{E(YY')} = \mathrm{E}\left[(\mathrm{X}\beta + e)(\mathrm{X}\beta + e)'\right]$$

$$= \sigma^2 tr(I-H) \qquad \mathrm{(I-H)X = X - X = 0}$$

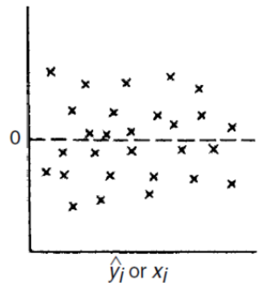
$$= \sigma^2 (n-p-1) \qquad \mathrm{tr(H) = p+1}$$

- $\mathrm{E}\left(\sum (Y_i - \bar{Y})^2\right) = E(Y'(I - J/n)Y) = ?$

- $\mathrm{E}\left(\sum (\hat{Y}_i - \bar{Y})^2\right) = E(Y'(H - J/n)Y) = ?$

8.1.3. Residuals when the model is correct

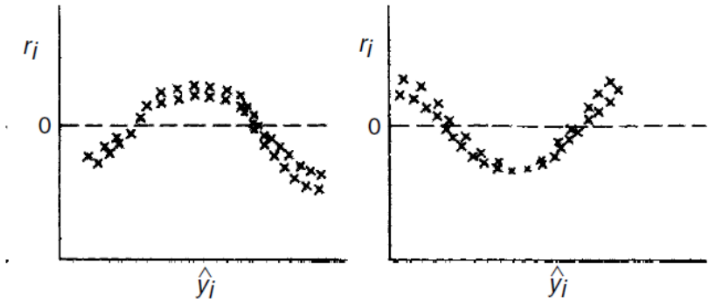
- Look at residual plot
  - y-axis = residuals
  - x-axis = term / combination of terms / fitted values
- Let U be any combinations of terms, we expect the Null Plot



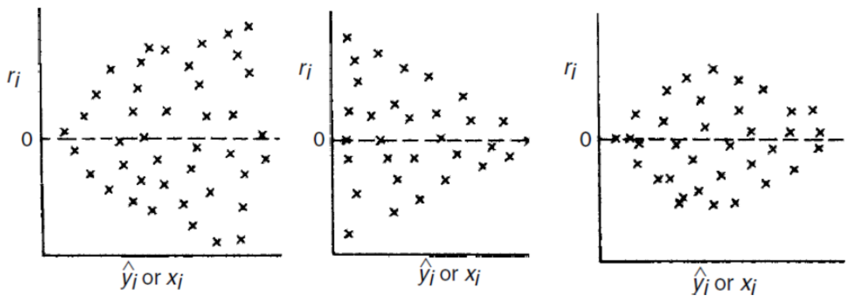
  - $E(\hat{e}_i \mid U) = 0$
    - mean level = 0
  - $Var(\hat{e}_i \mid U) = \sigma^2 (1 - h_{ii})$
    - variance is roughly constant (since h_{ii} are usually small)
    - The point with high leverage (h_{ii}~1) have small variance

8.1.4. Residuals when the model is NOT correct

- Residual plot far from the Null Plot
  - Mean level not 0



  - Variance not constant



  - Both