Stat 230 hw4. Yifan Zheng. 3035272865

1. Lec 9. Q1. Prove Lemma 1.

$$\sum_{i=1}^n (y_i-\bar{y})^2 = \sum_{i=1}^n (y_i-\hat{y}_i+\hat{y}_i-\bar{y})^2$$
$$= \sum_{i=1}^n (y_i-\hat{y})^2 + 2\sum_{i=1}^n (y_i-\hat{y}_i)(\hat{y}_i-\bar{y}) + \sum_{i=1}^n (\hat{y}_i-\bar{y})^2$$

Because $\sum_{i=1}^n (y_i-\hat{y}_i)(\hat{y}_i-\bar{y}) = \sum_{i=1}^n \varepsilon_i \hat{y}_i - \bar{y}\sum_{i=1}^n \varepsilon_i = 0$

Hence $\sum_{i=1}^n (y_i-\bar{y})^2 = \sum_{i=1}^n (y_i-\hat{y})^2 + \sum_{i=1}^n (\hat{y}_i-\bar{y})^2$

2. Lec 9. Q2. prove Theorem 1.

Because $\sum_{i=1}^n (y_i-\bar{y})(\hat{y}_i-\bar{y}) = \sum_{i=1}^n (\hat{y}_i-\bar{y}+y_i-\hat{y}_i)(\hat{y}_i-\bar{y})$
$= \sum_{i=1}^n (\hat{y}_i-\bar{y})^2 + \sum_{i=1}^n (y_i-\hat{y}_i)(\hat{y}_i-\bar{y}) = \sum_{i=1}^n (\hat{y}_i-\bar{y})^2$

Hence. $$\hat{\rho}^2_{\hat{y}y} = \frac{\left[\sum_{i=1}^n (\hat{y}_i-\bar{y})^2\right]^2}{\sum_{i=1}^n (y_i-\bar{y})^2 \cdot \sum_{i=1}^n (\hat{y}_i-\bar{y})^2} = \frac{\sum_{i=1}^n (\hat{y}_i-\bar{y})^2}{\sum_{i=1}^n (y_i-\bar{y})^2} = R^2$$

3. Lec 10. Q5 show $studr_i \sim t_{n-p-1}$

Proof: $\hat{\varepsilon}_{[-i]} = \hat{\varepsilon}_i/(1-h_{ii}) \sim N(0, (1-h_{ii})^{-2}\cdot\sigma^2(1-h_{ii}))$

$$\Rightarrow \frac{\hat{\varepsilon}_{[-i]}\cdot(1-h_{ii})}{\sqrt{\sigma^2}} \sim N(0,1).$$

$$\hat{\sigma}^2_{[-i]} = \frac{\tilde{\varepsilon}^T\tilde{\varepsilon}}{n-p-1} \quad \text{where } \tilde{\varepsilon} = (\tilde{\varepsilon}_1 \cdots \tilde{\varepsilon}_{i-1}, \tilde{\varepsilon}_{i+1}, \cdots \tilde{\varepsilon}_n)^T$$
$$= \frac{\tilde{\varepsilon}^T\tilde{\varepsilon}}{\sigma^2}\cdot\frac{\sigma^2}{n-p-1} \sim \frac{\sigma^2}{n-p-1}\cdot\chi^2_{n-p-1}$$

Hence $\dfrac{\hat{\varepsilon}_{[-i]}}{\sqrt{\hat{\sigma}^2_{[-i]}/(1-h_{ii})}} = \dfrac{\hat{\varepsilon}_{[-i]} \cdot \sqrt{1-h_{ii}}}{\sqrt{\hat{\sigma}^2_{[-i]}}} = \dfrac{\hat{\varepsilon}_{[-i]}\sqrt{1-h_{ii}}}{\sqrt{\sigma^2}} \Big/ \sqrt{\dfrac{\hat{\sigma}^2_{[-i]}}{\sigma^2}}$

Because $\dfrac{\hat{\varepsilon}_{[-i]}\sqrt{1-h_{ii}}}{\sqrt{\sigma^2}} \sim N(0,1)$, $\sqrt{\dfrac{\hat{\sigma}^2_{[-i]}}{\sigma^2}} \sim \dfrac{1}{n-p-1}\chi^2_{n-p-1}$

And they are independent. Hence $\text{standr}_i \sim t_{n-p-1}$

4. Lec 10. Q7. Prove $\text{cook}_i = \text{standr}_i^2 \times \dfrac{h_{ii}}{p(1-h_{ii})}$

Proof: $\text{standr}_i^2 \cdot \dfrac{h_{ii}}{p(1-h_{ii})} = \dfrac{\hat{\varepsilon}_i^T\hat{\varepsilon}_i \cdot h_{ii}}{(1-h_{ii})^2} \cdot \dfrac{1}{p\hat{\sigma}^2}$

$$
\begin{aligned}
\text{cook}_i &= (X\hat{\beta}_{[-i]} - X\hat{\beta})^T(X\hat{\beta}_{[-i]} - X\hat{\beta})/p\hat{\sigma}^2 \\
&= (1-h_{ii})^{-2}(X(X^TX)^{-1}x_i\hat{\varepsilon}_i)^T(X(X^TX)^{-1}x_i\hat{\varepsilon}_i)/p\hat{\sigma}^2 \\
&= (1-h_{ii})^{-2}\,\hat{\varepsilon}_i^T x_i^T(X^TX)^{-1}X^TX(X^TX)^{-1}x_i\hat{\varepsilon}_i/p\hat{\sigma}^2 \\
&= (1-h_{ii})^{-2} \cdot h_{ii} \cdot \varepsilon_i^T\varepsilon_i/p\hat{\sigma}^2
\end{aligned}
$$

Hence $\text{cook}_i = \text{standr}_i^2 \cdot \dfrac{h_{ii}}{p(1-h_{ii})}$

5. Lec11. Q8. The relationship between the standardized and studentized residual Show that

1. $(n-p-1)\hat{\sigma}^2_{[-i]} = (n-p)\hat{\sigma}^2 - \hat{\varepsilon}_i^2/(1-h_{ii})$,
2. there is a monotone relationship between the standardized and studentized residual:

$$\text{studr}_i = \text{standr}_i\sqrt{\frac{n-p-1}{n-p-\text{standr}_i^2}}.$$

(1). Proof: $\hat{\sigma}^2_{[-i]} = \dfrac{\tilde{\varepsilon}^T\tilde{\varepsilon}}{n-p-1}$ where $\tilde{\varepsilon} = (\tilde{\varepsilon}_1, \cdots \tilde{\varepsilon}_{i-1}\ \tilde{\varepsilon}_{i+1}, \cdots \tilde{\varepsilon}_n)^T$

where $\tilde{\varepsilon}_j = y_j - x_j\hat{\beta}_{[-i]}$, $j \neq i$,

Hence $(n-p-1)\hat{\sigma}^2_{[-i]} = (Y - X\hat{\beta}_{[-i]})^T(Y - X\hat{\beta}_{[-i]}) - \hat{\varepsilon}^2_{[-i]}$

$$= (Y - X\hat{\beta} + (1-h_{ii})^{-1}X(X^TX)^{-1}x_i\hat{\varepsilon}_i)^T(Y - X\hat{\beta} + (1-h_{ii})^{-1}X(X^TX)^{-1}x_i\hat{\varepsilon}_i) - \hat{\varepsilon}^2_{[-i]}$$

$$= \left(Y - X\hat{\beta} + \frac{\hat{\varepsilon}_i}{1-h_{ii}}\cdot\begin{bmatrix} h_{1i} \\ \vdots \\ h_{ni}\end{bmatrix}\right)^T\left(Y - X\hat{\beta} + \frac{\hat{\varepsilon}_i}{1-h_{ii}}\cdot\begin{bmatrix} h_{1i} \\ \vdots \\ h_{ni}\end{bmatrix}\right) - \frac{\hat{\varepsilon}_i^2}{(1-h_{ii})^2}$$

$$= (Y - X\hat{\beta})^T(Y - X\hat{\beta}) + \frac{\hat{\varepsilon}_i^2}{(1-h_{ii})^2}\cdot\sum_{j=1}^n h_{ji}^2 - \frac{\hat{\varepsilon}_i^2}{(1-h_{ii})^2}$$

$$= (n-p)\cdot\hat{\sigma}^2 - \frac{1-h_{ii}}{(1-h_{ii})^2}\hat{\varepsilon}_i^2 = (n-p)\hat{\sigma}^2 - \frac{\hat{\varepsilon}_i^2}{1-h_{ii}}$$

(2). $\text{studr}_i = \dfrac{\hat{\varepsilon}_{[-i]}}{\sqrt{\hat{\sigma}^2_{[-i]}/(1-h_{ii})}} = \dfrac{\hat{\varepsilon}_i}{1-h_{ii}}\cdot\dfrac{\sqrt{1-h_{ii}}\cdot\sqrt{n-p-1}}{\sqrt{(n-p)\hat{\sigma}^2 - \hat{\varepsilon}_i^2/(1-h_{ii})}} = \dfrac{\hat{\varepsilon}_i\sqrt{n-p-1}}{\sqrt{(1-h_{ii})(n-p)\hat{\sigma}^2 - \hat{\varepsilon}_i^2}}$

$\text{standr}_i = \dfrac{\hat{\varepsilon}_i}{\sqrt{\hat{\sigma}^2(1-h_{ii})}}$, $\text{standr}_i^2 = \dfrac{\hat{\varepsilon}_i^2}{\hat{\sigma}^2(1-h_{ii})}$

Hence $\text{standr}_i\cdot\sqrt{\dfrac{n-p-1}{n-p-\text{standr}_i^2}} = \dfrac{\hat{\varepsilon}_i\cdot\sqrt{n-p-1}}{\sqrt{\hat{\sigma}^2(1-h_{ii})(n-p) - \hat{\varepsilon}_i^2}}$

Hence $\text{studr}_i = \text{standr}_i\cdot\sqrt{\dfrac{n-p-1}{n-p-\text{standr}_i^2}}$

When standri increases. studri increases. there's monotone relationship.

6. Lec 11 Q 7.

7 Independence and correlation With scalar x and y, show that if $x \perp\!\!\!\perp y$, then $\rho_{yx} = 0$. With another variable w, if $x \perp\!\!\!\perp y \mid w$, does $\rho_{yx|w} = 0$ hold? If so, give a proof; otherwise, give a counterexample.

(1) if $X \perp\!\!\!\perp Y$. $E(XY) = E(X) \cdot E(Y)$.

Hence $Cov(X,Y) = E(XY) - E(X)E(Y) = 0$. then $\rho_{YX} = 0$.

(2). It does not hold.

Suppose $X, Y \overset{iid}{\sim} unif(0, w^2)$ and $w \sim N(0,1)$

$$\begin{aligned} Cov(X,Y) &= E(XY) - EXEY = E(E(XY|w)) - EXEY \\ &= E(E(X|w) \cdot E(Y|w)) - E(E(X|w)) E(E(Y|w)) \\ &= E(\tfrac{1}{2}w^2 \cdot \tfrac{1}{2}w^2) - E(\tfrac{1}{2}w^2) \cdot E(\tfrac{1}{2}w^2) \\ &= \tfrac{1}{4} Ew^4 - \tfrac{1}{4} Ew^2 Ew^2 = \tfrac{3}{4} - \tfrac{1}{4} = \tfrac{1}{2} \end{aligned}$$

$$\begin{aligned} Cov(X,w) &= E(Xw) - EXEw = E(E(Xw|w)) - E(E(X|w))Ew \\ &= E(\tfrac{1}{2}w^3) - E(\tfrac{1}{2}w^2) \cdot Ew = \tfrac{1}{2}Ew^3 - \tfrac{1}{2}Ew^2 Ew \\ &= 0 \end{aligned}$$

Hence. $Cov(X,Y) - Cov(X,w)Cov(Y,w) \cdot Var w \neq 0$

$\Rightarrow Cov(r_{X|w}, r_{Y|w}) \neq 0 \Rightarrow \rho_{X,Y|w} \neq 0$

7. Lec 11. Q9.

9 Best linear approximation of a cubic curve Assume that $x \sim N(0,1)$, $\varepsilon \sim N(0,\sigma^2)$, $x \perp\!\!\!\perp \varepsilon$, and $y = x^3 + \varepsilon$. Find the best linear approximation of y based on $(1, x)$.

Suppose best linear approximation is $\alpha + \beta \cdot X$.

$$(\alpha, \beta) = \underset{(a,b)}{\operatorname{argmin}}\ E(y - a - bX)^2,$$

where $\alpha = E(y) - E(X)\beta$ and. $\beta = \frac{Cov(X,y)}{Var(X)}$

Because $Cov(X,y) = Cov(X, X^3+\varepsilon) = Cov(X, X^3) = EX^4 = 3$

$Var(X) = 1$

$E(y) = E(X^3+\varepsilon) = E(X^3) + E(\varepsilon) = 0$

$E(X) = 0$

Hence $\alpha = 0$ $\beta = 3$

Hence best linear approximation is $3X$.