Lecture 7 Stat 331

Last class: Vector-matrix form of MLR model is given by: $\vec{Y} = X\vec{\beta} + \vec{\varepsilon}$

$$\vec{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{n\times 1} \quad X = \begin{array}{c} \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}_{n\times(p+1)} \\ \begin{array}{ccccc} \uparrow & \uparrow & \uparrow & & \uparrow \\ \vec{1} & \vec{x}_1 & \vec{x}_2 & \cdots & \vec{x}_p \end{array} \end{array} \quad \vec{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}_{(p+1)\times 1} \quad \vec{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n\times 1}$$

- Least-squares estimator of $\vec{\beta}$: $\hat{\vec{\beta}} = (X^TX)^{-1}X^T\vec{Y}$

  $\hat{\vec{\beta}} \sim MVN(\vec{\beta}, \sigma^2(X^TX)^{-1})$.

- In addition, $\hat{\beta}_j \sim N(\beta_j, \sigma^2 \underbrace{[(X^TX)]^{-1}_{jj}}_{(j,j)^{th} \text{ element of } (X^TX)^{-1}})$

$$\begin{bmatrix} (X^TX)^{-1}_{00} & \cdots & \cdots & \cdots \\ \vdots & (X^TX)^{-1}_{11} & \cdots & \vdots \\ \vdots & \cdots & \ddots & \vdots \\ \vdots & \cdots & \cdots & (X^TX)^{-1}_{pp} \end{bmatrix}$$

- Claim: $\hat{\sigma}^2 = SS(Res)/n-p-1$ is an unbiased estimator of σ^2. (Shown last class)

- Residuals (as a R.V.): $\vec{e} = \vec{Y} - \hat{\vec{\mu}} = \vec{Y} - X\hat{\vec{\beta}}$

  $= (I-H)\vec{Y}$ where $H = X(X^TX)^{-1}X^T$ (hat matrix).

  $\vec{e} \sim MVN(\vec{0}, \sigma^2(I-H))$.

Geometric Interpretation of LS.

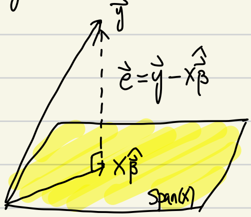
$$X = \begin{bmatrix} | & | & & | \\ \vec{1} & \vec{x}_1 & \cdots & \vec{x}_p \\ | & | & & | \end{bmatrix}$$

- $\text{Span}(X) = \underbrace{\{b_0\vec{1} + b_1\vec{x}_1 + \cdots + b_p\vec{x}_p}_{= X\vec{b} \text{ where } \vec{b} = (b_0, b_1, \ldots, b_p)^T} : b_0, \ldots, b_p \text{ are constants}\} \in \mathbb{R}^n$

  → linear combination of columns of X

- Span(X) represents all vector values given by $X\vec{b}$.

- $X\hat{\vec{\beta}} \in \text{Span}(X)$
- $\vec{y} \notin \text{Span}(X)$ ($\because \vec{\varepsilon}$ is variability not explained by X)



- In LS, $\hat{\vec{\beta}}$ is precisely a vector of estimates s.t. $X\hat{\vec{\beta}}$ is "closest" to $\vec{y}$.
- "closest" would be one where $\vec{e}$ are orthogonal to span(X) (i.e., $\vec{e}$ is orthogonal to each column of X.)

This implies:

(1) $\vec{1}^T(\vec{y}-X\hat{\vec{\beta}})=0 \iff \sum_{i=1}^{n} e_i = 0$

(2) $\vec{x}_j^T(\vec{y}-X\hat{\vec{\beta}})=0 \ (\forall j=1,\ldots,p) \iff \sum_{i=1}^{n} e_i x_{ij}=0 \ (\forall j)$

(3) $(X\hat{\vec{\beta}})^T(\vec{y}-X\hat{\vec{\beta}})=0 \iff \sum_{i=1}^{n} e_i \hat{\mu}_i = 0$ (exercise)

LHS of (1) and (2): $X^T\vec{e} = \begin{bmatrix} | & | & & | \\ \vec{1} & \vec{x}_1 & \cdots & \vec{x}_p \\ | & | & & | \end{bmatrix}^T \begin{bmatrix} | \\ \vec{e} \\ | \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} e_i \\ \sum_{i=1}^{n} e_i x_{i1} \\ \vdots \end{bmatrix}$

from lecture 6

$$X^T\vec{e} = X^T(\vec{y}-X\hat{\vec{\beta}}) = X^T(I-H)\vec{y}$$
$$= X^T\vec{y} - X^TX(X^TX)^{-1}X^T\vec{y} = X^T\vec{y}-X^T\vec{y}=0.$$

MLE: MLE (maximum likelihood estimator) $\hat{\theta} = \arg\max_{\theta} L(\theta|y)$

In MLR, $(Y_1,\ldots,Y_n)^T$ are independent. In addition $\vec{Y} \sim MVN(X\vec{\beta}, \sigma^2 I)$

The likelihood function is given by:

$$L(\vec{\beta},\sigma^2|\vec{y}) = \frac{1}{(2\pi)^{n/2}|\sigma^2 I|^{1/2}} \exp\left\{\frac{-1}{2\sigma^2}(\vec{y}-X\vec{\beta})^T(\vec{y}-X\vec{\beta})\right\}.$$

$$\ell(\vec{\beta},\sigma^2|\vec{y}) = \frac{-n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\underbrace{(\vec{y}-X\vec{\beta})^T(\vec{y}-X\vec{\beta})}_{S(\vec{\beta}) \text{ for LS.}}$$

$S(\vec{\beta},\sigma^2|\vec{y})=0$: (setting score to zero)

① $\dfrac{\partial \ell(\vec{\beta},\sigma^2|\vec{y})}{\partial\vec{\beta}}=0 \rightarrow \hat{\vec{\beta}}_{MLE} = (X^TX)^{-1}X^T\vec{Y}$

② $\dfrac{\partial \ell(\vec{\beta},\sigma^2|\vec{y})}{\partial\sigma^2}=0 \rightarrow \hat{\sigma}^2_{MLE} = \frac{1}{n}(\vec{Y}-X\hat{\vec{\beta}})^T(\vec{Y}-X\hat{\vec{\beta}})$

$= \frac{1}{n}\vec{e}^T\vec{e}$ (biased)

(Since $E(\vec{e}^T\vec{e}) = \sigma^2(n-p-1)$, last class)

Inference for MLR

Fact: $SS(Res)/\sigma^2 = \hat{\sigma}^2(n-p-1)/\sigma^2 \sim \chi^2_{n-p-1}$

Review: $\hat{\vec{\beta}} \sim MVN\left(\vec{\beta}, \sigma^2(X^TX)^{-1}\right)$

$$\hat{\beta}_j \sim N\left(\beta_j, \sigma^2 (X^TX)^{-1}_{jj}\right), \text{ or equivalently}$$

$$\frac{\hat{\beta}_j-\beta_j}{\sigma\sqrt{(X^TX)^{-1}_{jj}}} \sim N(0,1)$$

As before, σ^2 is not known so we replace it with $\hat{\sigma}^2 = SS(Res)/n-p-1$:

$$\frac{\hat{\beta}_j-\beta_j}{\hat{\sigma}\sqrt{(X^TX)^{-1}_{jj}}} \sim ?$$

$$\frac{(\hat{\beta}_j-\beta_j)/\sigma\sqrt{(X^TX)^{-1}_{jj}}}{\underbrace{\hat{\sigma}\sqrt{(X^TX)^{-1}_{jj}}/\sigma\sqrt{(X^TX)^{-1}_{jj}}}_{\sqrt{\hat{\sigma}^2/\sigma^2}}} = \frac{\overbrace{(\hat{\beta}_j-\beta_j)/\sigma\sqrt{(X^TX)^{-1}_{jj}}}^{N(0,1)}}{\sqrt{\underbrace{\frac{\hat{\sigma}^2(n-p-1)}{\sigma^2}}_{\chi^2_{n-p-1}}\cdot\frac{1}{n-p-1}}}$$

(Fact: $T = Z/\sqrt{U/k}$ where $Z \sim N(0,1)$ and $U \sim \chi^2_k$.
If $Z \perp\!\!\!\perp U$, $T \sim t_k$.)

Want to show $\hat{\beta}_j$ and $\hat{\sigma}^2 = \vec{e}^T\vec{e}/n-p-1$ are independent.
It suffices to show that $\hat{\beta}_j$ is independent of $\vec{e}$.

Pf. In order to show this, consider $(\hat{\vec{\beta}}, \vec{e})^T$

$$\begin{bmatrix}\hat{\vec{\beta}}\\ \vec{e}\end{bmatrix} = \begin{bmatrix}(X^TX)^{-1}X^T\vec{Y}\\ (I-H)\vec{Y}\end{bmatrix} = \begin{bmatrix}(X^TX)^{-1}X^T\\ I-H\end{bmatrix}\vec{Y} \quad \therefore \begin{bmatrix}\hat{\vec{\beta}}\\ \vec{e}\end{bmatrix} \text{ is also MVN.}$$

$$Var\begin{bmatrix}\hat{\vec{\beta}}\\ \vec{e}\end{bmatrix} = \begin{bmatrix}(X^TX)^{-1}X^T\\ I-H\end{bmatrix}\sigma^2 I\begin{bmatrix}(X^TX)^{-1}X^T\\ I-H\end{bmatrix}^T$$

$$= \sigma^2\begin{bmatrix}(X^TX)^{-1}X^T\\ I-H\end{bmatrix}\begin{bmatrix}\{(X^TX)^{-1}X^T\}^T & I-H\end{bmatrix}$$

$$= \sigma^2\begin{bmatrix}(X^TX)^{-1} & (X^TX)^{-1}X^T(I-H)\\ (I-H)X(X^TX)^{-1} & I-H\end{bmatrix}$$

• $(X^TX)^{-1}X^T(I-H) = (X^TX)^{-1}X^T - (X^TX)^{-1}X^T X(X^TX)^{-1}X^T = (X^TX)^{-1}X^T - (X^TX)^{-1}X^T = 0$

• $(I-H)X(X^TX)^{-1} = 0$

$$\text{Var}\begin{bmatrix}\hat{\vec{\beta}} \\ \vec{e}\end{bmatrix} = \sigma^2\begin{bmatrix}(X^TX)^{-1} & 0 \\ 0 & I-H\end{bmatrix}$$

- Since $(\hat{\vec{\beta}}, \vec{e})^T$ is MVN, by property 4 (lecture 5) of MVN, $\hat{\vec{\beta}}$ and $\vec{e}$ are independent.

$\Rightarrow \hat{\beta}_j$ is independent of $\vec{e}$. $\blacksquare$

$$\therefore \frac{\hat{\beta}_j - \beta_j}{\underbrace{\hat{\sigma}\sqrt{(X^TX)^{-1}_{jj}}}_{SE(\hat{\beta}_j)}} \sim t_{n-p-1}$$

For any β_j, we can find

1) $100(1-\alpha)\%$ CI : $\hat{\beta}_j \pm c\,SE(\hat{\beta}_j)$, $c = t_{1-\frac{\alpha}{2}, n-p-1}$ ($1-\frac{\alpha}{2}$ quantile of t_{n-p-1})
2) Test $H_0 : \beta_j = 0$ vs. $H_a : \beta_j \neq 0$.
   If $|t| = |\hat{\beta}_j / SE(\hat{\beta}_j)| > c$, then reject H_0.
   p-value $= 2P(T \geq |t|) < \alpha$, then reject H_0.

Inference for mean response

If we want to estimate mean response for $\vec{x}_c = (1, x_{c1}, \cdots, x_{cp})^T$ we can use :

$$\hat{\mu}_c = \hat{\beta}_0 + \hat{\beta}_1 x_{c1} + \cdots + \hat{\beta}_p x_{cp} = \vec{x}_c^T \hat{\vec{\beta}}$$

As a R.V. $E(\hat{\mu}_c) = \vec{x}_c^T E(\hat{\vec{\beta}})$

$= \vec{x}_c^T \vec{\beta}$

$\text{Var}(\hat{\mu}_c) = \vec{x}_c^T \text{var}(\hat{\vec{\beta}}) \vec{x}_c$

$= \sigma^2 \vec{x}_c^T (X^TX)^{-1} \vec{x}_c$ $\quad\therefore \hat{\mu}_c \sim N(\vec{x}_c^T\vec{\beta}, \sigma^2 \vec{x}_c^T (X^TX)^{-1}\vec{x}_c)$

$$\frac{\hat{\mu}_c - \vec{x}_c^T\vec{\beta}}{\underbrace{\hat{\sigma}\sqrt{\vec{x}_c^T (X^TX)^{-1}\vec{x}_c}}_{SE(\hat{\mu}_c)}} \sim t_{n-p-1}$$

$100(1-\alpha)\%$ CI is $\hat{\mu}_c \pm c\,\hat{\sigma}\sqrt{\vec{x}_c^T(X^TX)^{-1}\vec{x}_c}$, c is $1-\frac{\alpha}{2}$ quantile of t_{n-p-1} dist^n.

Prediction using MLR

For a new observation w/ explanatory variables $\vec{x}_0 = (1, x_{01}, \dots, x_{0p})^T$, we want to predict the response given $\vec{x}_0$. We can do this using:

$$\hat{y}_0 = \vec{x}_0^T \hat{\vec{\beta}}$$

Prediction error is given by $y_0 - \hat{y}_0$.

As a R.V., the properties of $Y_0 - \hat{Y}_0$ are:

① $E(Y_0 - \hat{Y}_0) = \vec{x}_0^T \vec{\beta} - \vec{x}_0^T E(\hat{\vec{\beta}}) = 0$

②
$$\begin{aligned} Var(Y_0 - \hat{Y}_0) &= Var(\varepsilon_0 - \vec{x}_0^T \hat{\vec{\beta}}) \\ &= Var(\varepsilon_0) + \vec{x}_0^T Var(\hat{\vec{\beta}}) \vec{x}_0 \\ &= \sigma^2 + \sigma^2 \vec{x}_0^T (X^TX)^{-1} \vec{x}_0 \\ &= \sigma^2 (1 + \vec{x}_0^T (X^TX)^{-1} \vec{x}_0). \end{aligned}$$

$$\frac{Y_0 - \hat{Y}_0}{\hat{\sigma}\sqrt{1 + \vec{x}_0^T (X^TX)^{-1} \vec{x}_0}} \sim t_{n-p-1}$$

$100(1-\alpha)\%$ PI is $\hat{y}_0 \pm c\,\hat{\sigma}\sqrt{1 + \vec{x}_0^T (X^TX)^{-1} \vec{x}_0}$

where c is $1-\frac{\alpha}{2}$ quantile of t_{n-p-1} dist^{n}.