Lecture 6 Stat 331

Last class: Random vector $\vec{Y} = (Y_1, \ldots, Y_n)^T$ where $Y_1, \ldots, Y_n$ are R.V.

- $E(\vec{Y}) = \left(E(Y_1), \ldots, E(Y_n)\right)^T = (\mu_1, \ldots, \mu_n)^T = \vec{\mu}_{n\times 1}$
- $Var(\vec{Y}) = \Sigma_{n\times n} = E\left[(\vec{Y} - E(\vec{Y}))(\vec{Y} - E(\vec{Y}))^T\right]$
  - → Variance-covariance matrix of $\vec{Y}$ is symmetric, positive semi-definite.
  - → If $Y_1, \ldots, Y_n$ are independent $\Rightarrow \Sigma$ is a diagonal matrix

Multiple Linear Regression model

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i, \quad \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

- $(x_{i1}, \ldots, x_{ip})$: explanatory variables for observation i.
- $\beta_j\ (j = 1, \ldots, p)$: regression coefficient or parameter representing the association between x_j and response.

We can write the model in vector-matrix form:

$$\underbrace{\vec{Y}}_{\text{random vector}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{n\times 1} = \underbrace{\begin{bmatrix} \beta_0 + \beta_1 x_{11} + \cdots + \beta_p x_{1p} \\ \beta_0 + \beta_1 x_{21} + \cdots + \beta_p x_{2p} \\ \vdots \\ \beta_0 + \beta_1 x_{n1} + \cdots + \beta_p x_{np} \end{bmatrix}_{n\times 1}}_{(*)} + \overbrace{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n\times 1}}^{\vec{\varepsilon}\text{ : random vector.}}$$

Moreover, (*) has a matrix form given by $X\vec{\beta}$ where

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}_{n\times(p+1)} ; \quad \vec{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}_{(p+1)\times 1}$$

where X is the design matrix

$$\therefore \vec{Y} = X\vec{\beta} + \vec{\varepsilon}$$

Interpretation of $\vec{\beta}$:

β_0 : average or mean response when $x_1 = x_2 = \cdots = x_p = 0$.

$\beta_j\ (j = 1, \ldots, p)$: change in mean response for a unit increase in x_j, holding all other covariates constant.

by ppty (3) of MVN

Note: Since $\varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$, then random error vector $\vec{\varepsilon} \sim MVN(\vec{0}, \sigma^2 I)$

where $I_{n\times n}$ identity matrix $= \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}$

$\therefore \vec{Y} \sim MVN(X\vec{\beta}, \sigma^2 I)$; $\vec{\mu} = X\vec{\beta}$.

Least-square estimation : Analogous to SLR, we want to minimize $\vec{\beta}$ over an objective function :

$$\underset{\vec{\beta}}{\operatorname{argmin}} \underbrace{\sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip})\right)^2}_{S(\vec{\beta})}$$

In matrix form, this is equivalent to

$$\underset{\vec{\beta}}{\operatorname{argmin}} \left[\underbrace{(\vec{y} - X\vec{\beta})^T(\vec{y} - X\vec{\beta})}_{S(\vec{\beta})}\right]$$

Then taking derivative wrt $\vec{\beta}$, we get

$$\frac{\partial S(\vec{\beta})}{\partial \vec{\beta}} = \frac{\partial}{\partial \beta}\left[(\vec{y} - X\vec{\beta})^T(\vec{y} - X\vec{\beta})\right] = \frac{\partial}{\partial \beta}\left[\vec{y}^T y - \underbrace{\vec{y}^T X}_{\vec{a}^T}\vec{\beta} - \vec{\beta}^T \underbrace{X^T \vec{y}}_{\vec{a}} + \vec{\beta}^T \underbrace{X^T X}_{A}\vec{\beta}\right]$$

Some useful matrix derivatives (scalar differentiation wrt vector):

- If $z = \vec{a}^T\vec{y}$ where $\vec{a}, \vec{y}$ are $n\times 1$ vectors, then $\frac{\partial z}{\partial \vec{y}} = \vec{a}$
- If $z = \vec{y}^T\vec{a}$, then $\frac{\partial z}{\partial \vec{y}} = \vec{a}$
- If A is $n\times n$ matrix and let $z = \vec{y}^T A\vec{y}$, then $\frac{\partial z}{\partial \vec{y}} = A\vec{y} + A^T\vec{y}$

Moreover, if A is symmetric, then $\frac{\partial z}{\partial \vec{y}} = 2A\vec{y}$.

$$\frac{\partial S(\vec{\beta})}{\partial \vec{\beta}} = -X^T\vec{y} - X^T\vec{y} + 2X^TX\vec{\beta} \quad (X^TX \text{ is symmetric}).$$

$$= 0$$

$$X^TX\vec{\beta} = X^T\vec{y} \quad \left(\Rightarrow X^T(\vec{y} - X\vec{\beta}) = 0\right).$$

$$\vec{\beta} = (X^TX)^{-1}X^T\vec{y}$$

$\therefore$ Least-squares (LS) estimator of $\vec{\beta}$ is given by $\hat{\vec{\beta}} = (X^TX)^{-1}X^T\vec{Y}$

Given a data sample, the LS estimate of $\vec{\beta}$ is $\hat{\vec{\beta}} = (X^TX)^{-1}X^T\vec{y}$

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

Note that $\vec{Y}$ is $MVN(X\vec{\beta}, \sigma^2 I)$ and that $(X^TX)^{-1}X^T$ is a $(p+1)\times n$ matrix of constants. By property (1) of MVN distⁿ, we know that $\hat{\vec{\beta}}$ is MVN distributed.

Assuming that X^TX is invertible (full rank is $p+1$; columns of X are linearly indpt.).

① $E(\hat{\vec{\beta}}) = \vec{\beta}$

Pf: $E(\hat{\vec{\beta}}) = E\left((X^TX)^{-1}X^T\vec{Y}\right)$

$$= (X^TX)^{-1}X^T E(\vec{Y}) = \underbrace{(X^TX)^{-1}X^TX}_{I}\vec{\beta} = \vec{\beta} \quad \blacksquare$$

② $Var(\hat{\vec{\beta}}) = Var\left((X^TX)^{-1}X^T\vec{Y}\right)$

$$= (X^TX)^{-1}X^T Var(\vec{Y}) X (X^TX)^{-1} \quad (\because X^TX \text{ is symmetric})$$

$$= \sigma^2 (X^TX)^{-1}X^TX(X^TX)^{-1}$$

$$= \sigma^2 (X^TX)^{-1} \quad \blacksquare$$

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

③ $\hat{\beta}_j \sim N\left(\beta_j, \sigma^2 \underbrace{\left[(X^TX)^{-1}\right]_{jj}}_{\text{diagonal of } (X^TX)^{-1} \text{ corresponding to } \hat{\beta}_j}\right)$ (by marginal normality ppty (2) of MVN dist).

Estimation of σ^2

$\vec{e} = \vec{y} - X\hat{\vec{\beta}}$

As before, we define residual $e_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_p x_{ip})$

We define $SS(Res) = \sum_{i=1}^{n} e_i^2$ and we will use $SS(Res)$ to estimate variance of random error, σ^2.

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

In $X\hat{\vec{\beta}} = \underbrace{X(X^TX)^{-1}X^T}_{H_{n\times n}}\vec{Y}$ let $H = X(X^TX)^{-1}X^T$ where H is hat matrix.

$(H = X(X^TX)^{-1}X^T)$ Properties of H:

(1) H is symmetric : $H = H^T$

(2) H is idempotent : $H = HH$

(3) $(I-H)$ is symmetric

(4) $(I-H)$ is idempotent.

Verify (1): $H^T = [X(X^TX)^{-1}X^T]^T$
$= X[X(X^TX)^{-1}]^T = X(X^TX)^{-1}X^T = H.$

(2): $HH = X(X^TX)^{-1}X^TX(X^TX)^{-1}X^T$
$= X(X^TX)^{-1}X^T = H.$

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

$$E\left(\sum_{i=1}^n e_i^2\right) = E(\vec{e}^T\vec{e}) = E[\vec{Y}^T(I-H)(I-H)\vec{Y}]$$

(b/c $\vec{e} = \vec{Y} - X\hat{\vec{\beta}} = \vec{Y} - H\vec{Y} = (I-H)\vec{Y}$).

$$= E[\vec{Y}^T(I-H)\vec{Y}]$$

$$= E[tr(\underbrace{\vec{Y}^T}_{A}\underbrace{(I-H)\vec{Y}}_{B})]$$ (b/c $\vec{Y}^T(I-H)\vec{Y}$ is scalar).

$$= E[tr((I-H)\vec{Y}\vec{Y}^T)]$$ $\because tr(AB) = tr(BA)$

$$= tr[(I-H)E(\vec{Y}\vec{Y}^T)]$$ $\because E(tr(A)) = tr\,E(A)$.

$$\left(\begin{aligned} E(\vec{Y}\vec{Y}^T) &= E[(X\vec{\beta}+\vec{\varepsilon})(X\vec{\beta}+\vec{\varepsilon})^T]. \\ &= X\vec{\beta}\vec{\beta}^TX^T + E(\vec{\varepsilon}\vec{\varepsilon}^T) \\ &= X\vec{\beta}\vec{\beta}^TX^T + \sigma^2 I \end{aligned}\right)$$

$$\therefore E\left(\sum_{i=1}^n e_i^2\right) = tr((I-H)(X\vec{\beta}\vec{\beta}^TX^T + \sigma^2 I))$$

$$= tr((I-H)\sigma^2 I)$$ $\because (I-H)X = 0$

$$= \sigma^2 tr(I-H)$$. (here $I_{n\times n}$ matrix) $tr(A+B) = tr(A) + tr(B)$

$$= \sigma^2(tr(I) - tr(H)) = \sigma^2(n - tr(H))$$

$\downarrow$ $tr(H) = tr(X(X^TX)^{-1}X^T) = tr(X^TX(X^TX)^{-1}) = tr(I_{(p+1)\times(p+1)}) = p+1.$

$$= \sigma^2(n-p-1)$$

$\therefore \hat{\sigma}^2 = SS(Res)/n-p-1$ is unbiased estimator of σ^2. ▨

Properties of residuals (as a R.V).

$$\vec{e} = \vec{Y} - X\hat{\vec{\beta}} = (I-H)\vec{Y}$$

$\because$ $\vec{e}$ is a linear transformation of $\vec{Y}$ and since $\vec{Y}$ is MVN distributed

$\therefore$ $\vec{e}$ is MVN.

① $E(\vec{e}) = E[(I-H)\vec{Y}]$.

$= (I-H)E(\vec{Y})$

$= (I-H)X\vec{\beta} = X\vec{\beta} - HX\vec{\beta} = X\vec{\beta} - X(X^TX)^{-1}X^TX\vec{\beta}$

$= X\vec{\beta} - X\vec{\beta} = \vec{0}$

② $Var(\vec{e}) = Var[(I-H)\vec{Y}]$

$= (I-H)\,Var(\vec{Y})\,(I-H)$

$= (I-H)\,\sigma^2 I\,(I-H)$.

$= \sigma^2(I-H)$,

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