lec5 : Stat 331

Last class In SLR , $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

$Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$ (independent)

Linearity
Independence (ε)
Normality (ε)
Equal variance (ε)

- $100(1-\alpha)\%$ CI for β_1 is $\hat{\beta}_1 \pm c\,SE(\hat{\beta}_1)$, $c = t_{1-\frac{\alpha}{2}, n-2}$ ($1-\frac{\alpha}{2}$ quantile of t_{n-2} distⁿ).

(analogously 95% CI for β_0 is $\hat{\beta}_0 \pm c\,SE(\hat{\beta}_0)$)

$SE(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2 / S_{xx}}$ $\left(SE(\hat{\beta}_0) = \sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{\bar{x}}{S_{xx}}\right)} \right)$

- Fitted value (mean response) for $x = x_p$: $\hat{\mu}_p = \hat{\beta}_0 + \hat{\beta}_1 x_p$.

Sampling dist^n : $\frac{\hat{\mu}_p - \mu_p}{SE(\hat{\mu}_p)} \sim t_{n-2}$, $SE(\hat{\mu}_p) = \sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{(x_p - \bar{x})^2}{S_{xx}}\right)}$

$100(1-\alpha)\%$ CI : $\hat{\mu}_p \pm c\,SE(\hat{\mu}_p)$

- Predict a response for new observation $x = x_0$: $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$.

Sampling dist^n : $\frac{Y_0 - \hat{Y}_0}{SE(Y_0 - \hat{Y}_0)} \sim t_{n-2}$, $SE(Y_0 - \hat{Y}_0) = \sqrt{\hat{\sigma}^2\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)}$.

prediction error

$100(1-\alpha)\%$ PI : $\hat{y}_0 \pm c\,SE(y_0 - \hat{y}_0)$.

Intro into Multiple Linear Regression (MLR):

In SLR, we studied the relationship between a response variable and a single explanatory variable. In MLR, we study its relationship with multiple explanatory variables.

Data : n data points / observations. Instead of pairs (x_i, y_i) in SLR, we have $p+1$-tuple: consisting of a response and vector of covariates : $(y_i, x_{i1}, \dots, x_{ip})$.

x_{ij} $(j=1,\dots,p)$: j^{th} explanatory variable of i^{th} observation.

Def^{n}: A random vector → For a set of R.V. $(Y_1, \dots, Y_n)$:

$$\vec{Y} = (Y_1, \dots, Y_n)^T \text{ is a random vector.}$$

- Mean of $\vec{Y} = E(\vec{Y}) = (E(Y_1), \dots, E(Y_n))^T = (\mu_1, \dots, \mu_n)^T = \vec{\mu}$
  (Here, $\vec{Y}$ and $\vec{\mu}$ are $n \times 1$ vectors)
- Variance - Covariance matrix of $\vec{Y}$ (or Σ) $= E\left[(\vec{Y}-E(\vec{Y}))(\vec{Y}-E(\vec{Y}))^T\right]$

$$\Sigma_{n\times n} = Var(\vec{Y}) = \begin{bmatrix} Var(Y_1) & Cov(Y_1,Y_2) & \dots & Cov(Y_1,Y_n) \\ & Var(Y_2) & \dots & Cov(Y_2,Y_n) \\ & & & \vdots \\ & & & Var(Y_n) \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ & \sigma_2^2 & \dots & \sigma_{2n} \\ & & \ddots & \vdots \\ & & & \sigma_n^2 \end{bmatrix}$$

Properties of Σ: (1) Symmetric matrix $(\Sigma = \Sigma^T ; \Sigma_{ij} = \Sigma_{ji})$

(2) Positive semi-definite : $\vec{a}^T \Sigma \vec{a} \geq 0$, $\forall \vec{a}_{n\times 1} \in R^n$

(3) If $(Y_1, \dots, Y_n)$ are independent, then Σ is a diagonal matrix (converse is not necessarily true, exception when $\vec{Y}$ is MVN, seen later)

Properties of random vector:

Let $\vec{a}$ be a $n\times 1$ vector of constants $(a_1, \dots, a_n)^T$

Let A be a $p\times n$ matrix of constants

(1) $E(\vec{a}^T\vec{Y}) = \vec{a}^T E(\vec{Y})$

(2) $E(A\vec{Y}) = A E(\vec{Y})$

(3) $Var(\vec{a}^T\vec{Y}) = \vec{a}^T Var(\vec{Y})\vec{a}$

(4) $Var(A\vec{Y}) = A\, Var(\vec{Y}) A^T$

Pf (4): $Var(A\vec{Y}) = E\left[(A\vec{Y}-E(A\vec{Y}))(A\vec{Y}-E(A\vec{Y}))^T\right]$
$= E\left[A(\vec{Y}-E(\vec{Y}))\{A(\vec{Y}-E(\vec{Y}))\}^T\right]$ $\quad (AB)^T = B^TA^T$
$= E\left[A(\vec{Y}-E(\vec{Y}))(\vec{Y}-E(\vec{Y}))^TA^T\right]$
$= A\,E\left[(\vec{Y}-E(\vec{Y}))(\vec{Y}-E(\vec{Y}))^T\right]A^T$
$= A\,Var(\vec{Y})A^T$ ∎

Q: $\vec{Y} = (Y_1, Y_2, Y_3)^T$, $E(\vec{Y}) = \begin{bmatrix}1\\2\\3\end{bmatrix}$, $Var(\vec{Y}) = \begin{bmatrix}2 & -1 & 0\\ -1 & 2 & -1\\ 0 & -1 & 2\end{bmatrix}$, $A = \begin{bmatrix}1&2&3\\4&5&6\\7&8&9\end{bmatrix}$, $\vec{a} = \begin{bmatrix}0\\1\\2\end{bmatrix}$

(1) $E(\vec{a}^T\vec{Y})$, (2) $Var(\vec{a}^T\vec{Y})$, (3) $E(A\vec{Y})$, (4) $Var(A\vec{Y})$

A: (1) $E(\vec{a}^T\vec{Y}) = [0\ 1\ 2]\begin{bmatrix}1\\2\\3\end{bmatrix} = 8$ $\quad$ (3) $\begin{bmatrix}14\\32\\50\end{bmatrix}$ $\quad$ (4) $\begin{bmatrix}12 & 24 & 36\\ 24 & 54 & 84\\ 36 & 84 & 132\end{bmatrix}$

(2) $Var(\vec{a}^T\vec{Y}) = [0\ 1\ 2]\,Var(\vec{Y})\begin{bmatrix}0\\1\\2\end{bmatrix} = 6$

Multivariate Normal Dist^n (MVN)

$\vec{Y} = (Y_1, \dots, Y_n)^T$ is MVN distributed if pdf is given by

$$f(\vec{y}; \vec{\mu}, \Sigma) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}\underbrace{(\vec{y}-\vec{\mu})^T}_{1\times n}\underbrace{\Sigma^{-1}}_{n\times n}\underbrace{(\vec{y}-\vec{\mu})}_{n\times 1}\right\}$$

here Σ^{-1} is the inverse variance-covariance matrix of $\vec{Y}$

(analogous to the one-dimensional case)

$E(\vec{Y}) = \vec{\mu}_{n\times 1}$, $Var(\vec{Y}) = \Sigma_{n\times n}$ $\quad \therefore \vec{Y} \sim MVN(\vec{\mu}, \Sigma)$.

Properties of MVN distribution

Suppose that $\vec{Y} \sim MVN(\vec{\mu}, \Sigma)$ and let $\vec{a}$ be a $n\times 1$ vector of constants and let A be $p\times n$ matrix of constants.

(1) Transformation invariant (linear transformation of MVN is also Normal)

$$\vec{a}^T\vec{Y} \sim N(\vec{a}^T\vec{\mu}, \vec{a}^T\Sigma\vec{a})$$
$$A\vec{Y} \sim MVN(A\vec{\mu}, A\Sigma A^T)$$

(If $\vec{b}$ is a $p\times 1$ vector of constants, $A\vec{Y}+\vec{b} \sim MVN(A\vec{\mu}+\vec{b}, A\Sigma A^T)$)

(2) Marginal dist^n of $\vec{Y}$ is Normal (marginal Normality).

$\vec{Y} \sim MVN(\vec{\mu}, \Sigma)$, then each element of $\vec{Y}$ is Normal: $Y_i \sim N(\mu_i, \Sigma_{ii})$

where μ_i: i^{th} element of $\vec{\mu}$

Σ_{ii}: i^{th} diagonal of Σ $(= \sigma_i^2)$

In general, for any partition of $\vec{Y}$:

$$\vec{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_k \\ Y_{k+1} \\ \vdots \\ Y_n \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix} Y_1 \\ \vdots \\ Y_k \end{matrix}}\right\} \vec{Y}_1 \\ \left.\vphantom{\begin{matrix} Y_1 \\ \vdots \\ Y_k \end{matrix}}\right\} \vec{Y}_2 \end{matrix} = \begin{bmatrix} \vec{Y}_1 \\ \vec{Y}_2 \end{bmatrix} \; ; \; \vec{\mu} = \begin{bmatrix} \vec{\mu}_1 \\ \vec{\mu}_2 \end{bmatrix} \; ; \; \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

then $\vec{Y}_1 \sim MVN(\vec{\mu}_1, \Sigma_{11})$; $\vec{Y}_2 \sim MVN(\vec{\mu}_2, \Sigma_{22})$

(3) If $Y_i \sim N(\mu_i, \sigma_i^2)$ and $(Y_1, \dots, Y_n)^T$ are independent, then
$\vec{Y} = (Y_1, \dots, Y_n)^T \sim MVN(\vec{\mu}, \Sigma)$ where $\Sigma = diag(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2)$

(4) For MVN random vector $\vec{Y}$, $Var(\vec{Y})$ diagonal matrix $\Longleftrightarrow$ Y_i's are independent
(i.e. when $Cov(Y_i, Y_j) = 0 \ \forall i \neq j$ then $Y_i \perp\!\!\!\perp Y_j$)
General case: If $\begin{bmatrix} \vec{Y}_1 \\ \vec{Y}_2 \end{bmatrix}$ is MVN, they are independent iff $Cov(\vec{Y}_1, \vec{Y}_2) = 0$.

(5) If $\vec{Y} \sim MVN(\vec{\mu}, \Sigma)$, let $\vec{V} = A\vec{Y}$ and $\vec{W} = B\vec{Y}$ (for A, B matrices of constants)
then $\vec{V}$ and $\vec{W}$ are indpt iff $A\Sigma B^T = 0$.
Pf. $\Rightarrow$ If $\vec{V}$ and $\vec{W}$ are indpt, then $A\Sigma B^T = 0$

$Cov(\vec{V}, \vec{W}) = 0$ (by indpt)

$Cov(A\vec{Y}, B\vec{Y}) = 0 \Rightarrow A\,Cov(\vec{Y}, \vec{Y})B^T = 0 \Rightarrow A\Sigma B^T = 0$

$\Leftarrow$ If $A\Sigma B^T = 0$ then $\vec{V}$ and $\vec{W}$ are indpt.

By (1), $\begin{bmatrix} \vec{V} \\ \vec{W} \end{bmatrix}$ is MVN.

Then (4) $\vec{V}$ and $\vec{W}$ are independent.