Lecture 6 (part 2)

RECAP OF THE MULTIPLE LINEAR MODEL

we observe $(y_i, x_{i1}, \dots, x_{ip})$ for $i = 1, \dots, n$

- MODEL SPECIFICATION

$$Y_i = \mu_i + \varepsilon_i = \beta_1 \overset{\overset{1\ \forall i \text{ if intercept}}{\parallel}}{x_{i1}} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i$$

$$\Rightarrow \underline{Y} = X\underline{\beta} + \underline{\varepsilon}$$

- ASSUMPTIONS
  1. normality, homoscedasticity, independence
     $\underline{\varepsilon} \sim N_n(\underline{0}, \sigma^2 I_n)$
     or, equivalently $\underline{Y} \sim N_n(X\underline{\beta}, \sigma^2 I_n)$
  2. linearity $\underline{\mu} = X\underline{\beta}$
  3. absence of multicollinearity of the covariates
     or, equivalently, $\text{rank}(X) = p$ (X has full rank)

- ESTIMATION
  - the MLE of $\underline{\beta}$ is $\hat{\underline{\beta}} = (X^T X)^{-1} X^T \underline{y}$
  - the predicted values
    $$\hat{\underline{y}} = X\hat{\underline{\beta}} = X \cdot (X^T X)^{-1} X^T \underline{y} = P \cdot \underline{y} = \hat{\underline{\mu}}$$
    with $P = X(X^T X)^{-1} X^T$ projection matrix on $\mathcal{C}(X)$
    $\Rightarrow P$ $n \times n$ symmetric, idempotent, rank $= p$
  - the residuals
    $$\underline{e} = \underline{y} - \hat{\underline{y}} = \underline{y} - X\hat{\underline{\beta}} = \underline{y} - X(X^T X)^{-1} X^T \underline{y} = (I_n - X(X^T X)^{-1} X^T)\underline{y} = (I_n - P)\underline{y}$$
    $(I_n - P)$ projection matrix on the subspace of $\mathbb{R}^n$ orthogonal to $\mathcal{C}(X)$
    $\Rightarrow (I_n - P)$ $n \times n$ symmetric, idempotent, rank $= n - p$
  - the MLE estimate of σ^2 is
    $$\hat{\sigma}^2 = \frac{1}{n} \underline{e}^T \underline{e} = \frac{1}{n} (\underline{y} - X\hat{\underline{\beta}})^T (\underline{y} - X\hat{\underline{\beta}})$$

PROPERTIES OF THE ESTIMATORS

Given $\underline{y} = (y_1, \dots, y_n)$ and X $(n \times p)$
we have seen that the estimates are
$$\hat{\underline{\beta}} = (X^T X)^{-1} X^T \underline{y}$$
$$\hat{\sigma}^2 = \frac{1}{n} \underline{e}^T \underline{e}$$
$$s^2 = \frac{1}{n-p} \underline{e}^T \underline{e}$$
we want to derive their exact distribution to perform inference.

- DISTRIBUTION OF $\hat{\underline{\beta}}(\underline{Y})$

estimator $\hat{\underline{\beta}}(Y) = (X^T X)^{-1} X^T \underline{Y}$ linear in $\underline{Y}$ (i.e. $\hat{\underline{\beta}}(Y) = A \cdot \underline{Y}$) with $\underline{Y} \sim N_n(X\underline{\beta}, \sigma^2 I_n)$

- LINEAR TRANSFORMATIONS of MULTIVARIATE GAUSSIAN RANDOM VECTORS

$\underline{z} \sim N_d(\underline{\mu}, \Sigma)$, A $(k \times d)$ matrix, $\underline{b} \in \mathbb{R}^k$

$$\Rightarrow T = \underset{(k\times d)}{A}\,\underset{(d\times 1)}{\underline{z}} + \underset{(k\times 1)}{\underline{b}} \sim N_k(A\underline{\mu} + \underline{b}, A\Sigma A^T)$$

$$\mathbb{E}[\underline{T}] = \mathbb{E}[A\underline{z} + \underline{b}] = A\,\mathbb{E}[\underline{z}] + \underline{b} = A\underline{\mu} + \underline{b}$$

$$\text{var}(\underline{T}) = \text{var}(A\underline{z} + b) = \text{var}(A\underline{z}) = \mathbb{E}[(Az - \mathbb{E}[Az])(Az - \mathbb{E}[Az])^T]$$
$$= \mathbb{E}[(Az - A\mu)(Az - A\mu)^T] = \mathbb{E}[A(z - \mu)(z - \mu)^T A^T] = A \Sigma A^T$$

$\Rightarrow \hat{\underline{\beta}}(\underline{Y}) = A\underline{Y}$ with $A = (X^T X)^{-1} X^T$ $p \times n$ rank $= p$

$$\mathbb{E}[\hat{\underline{\beta}}(\underline{Y})] = A\underline{\mu} = (X^T X)^{-1} X^T X \underline{\beta} = \underline{\beta}$$

$$\text{var}(\hat{\underline{\beta}}(\underline{Y})) = A\Sigma A^T = A(\sigma^2 I_n)A^T = \sigma^2 AA^T = \sigma^2 \cancel{(X^T X)^{-1}} \cancel{X^T} X((X^T X)^{-1})^T = \sigma^2 (X^T X)^{-1}$$

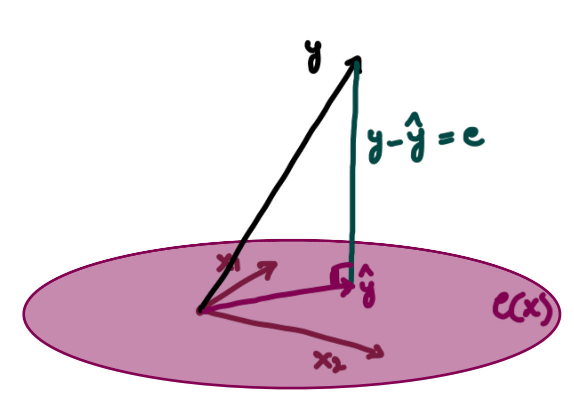
$$\Rightarrow \hat{\underline{\beta}}(\underline{Y}) \sim N_p(\underline{\beta}; \sigma^2 (X^T X)^{-1})$$

the marginal is $\hat{\beta}_j(\underline{Y}) \sim N_1(\beta_j; \sigma^2 [(X^T X)^{-1}]_{(j,j)})$

and the covariance is $\text{cov}(\hat{\beta}_j(Y), \hat{\beta}_s(Y)) = \sigma^2 [(X^T X)^{-1}]_{(j,s)}$

Notice that the variance is $\sigma^2 (X^T X)^{-1}$: once again, we need linearly independent $(\underline{x}_1, \dots, \underline{x}_p)$.
indeed $(X^T X)^{-1} = \frac{1}{\det(X^T X)} \cdot [\dots]$
If they are collinear $\det(X^T X) = 0$ and it is not invertible
However, if they are almost collinear $(\det(X^T X) \approx 0)$ the variance of the estimator explodes
(not good)

- DISTRIBUTION OF THE RESIDUALS



$$\underline{e} = \underline{y} - \hat{\underline{y}} = (I_n - P)\underline{y}$$

projection of $\underline{y}$ onto the subspace of $\mathbb{R}^n$ perpendicular to $\mathcal{C}(X)$
Let' study the corresponding random quantity $\underline{E}$

$$\underline{E} = (I_n - P)\underline{Y}$$
$$= (I_n - P)(X\underline{\beta} + \underline{\varepsilon})$$
$$= \underbrace{(I_n - P)X\underline{\beta}}_{=0} + (I_n - P)\underline{\varepsilon} = (I_n - P)\underline{\varepsilon} \qquad \underline{\varepsilon} \sim N(\underline{0}, \sigma^2 I_n)$$

$$(I_n - P)X\underline{\beta} = X\underline{\beta} - PX\underline{\beta} = \underline{0}$$

indeed, $(I_n - P)X$ is the projection of X on the space $\perp \mathcal{C}(X)$

$$PX = X(X^T X)^{-1} X^T X = X$$

Hence, $\underline{E} = (I_n - P)\underline{\varepsilon}$ $\quad \underline{\varepsilon} \sim N(\underline{0}, \sigma^2 I_n)$
linear combination of a Gaussian is Gaussian $\rightarrow \underline{E} \sim N$

$$\mathbb{E}[\underline{E}] = (I_n - P)\underline{0} = \underline{0}$$
$$\text{var}(\underline{E}) = (I_n - P)\,\text{var}(\underline{\varepsilon})\,(I_n - P)^T = \sigma^2 (I_n - P)(I_n - P)^T = \sigma^2 (I_n - P)$$
$$\Rightarrow \underline{E} \sim N_n(\underline{0}, (I_n - P)\sigma^2)$$

(i.e. $\text{var}(E_i) = \sigma^2 (I_n - P)_{(ii)} \rightarrow$ not homoscedastic)

- DISTRIBUTION OF $\hat{\sigma}^2(\underline{Y})$

$$\hat{\sigma}^2(\underline{Y}) = \frac{\underline{E}^T \underline{E}}{n} \quad \text{and} \quad \frac{\underline{E}^T \underline{E}}{\sigma^2} = \frac{n \hat{\sigma}^2(Y)}{\sigma^2} \sim \chi^2_{n-p}$$

$$\Rightarrow \mathbb{E}[\hat{\sigma}^2] = \frac{\sigma^2}{n}(n - p)$$

$n - p = \text{rank}(I_n - P)$
$= \#\text{units} - \#\text{covariates}$

As usual, we obtain an unbiased estimator as

$$\hat{S}^2 = \frac{\underline{E}^T \underline{E}}{n-p} = \frac{n\hat{\sigma}^2}{n-p} \quad \text{with} \quad \frac{(n-p)\hat{S}^2}{\sigma^2} \sim \chi^2_{n-p} \quad \Rightarrow \mathbb{E}[\hat{S}^2] = \sigma^2$$

moreover,
$$\hat{\underline{\beta}}(\underline{Y}) \perp\!\!\!\perp \hat{\sigma}^2(\underline{Y})$$
$$\hat{\underline{\beta}}(\underline{Y}) \perp\!\!\!\perp \hat{S}^2$$