DISTRIBUTION OF THE ESTIMATORS

Given $\underline{Y} = (Y_1, \dots, Y_n)$ and X $(n \times p)$

we have seen that the estimators are

$$\hat{\underline{\beta}} = (X^TX)^{-1}X^T\underline{Y}$$
$$\hat{\Sigma}^2 = \frac{1}{n}\underline{E}^T\underline{E} \qquad \text{with } \underline{E} = \underline{Y} - \hat{\underline{Y}} = \underline{Y} - X\hat{\underline{\beta}}$$
$$S^2 = \frac{1}{n-p}\underline{E}^T\underline{E}$$

we want to derive their exact distribution to perform inference.

• DISTRIBUTION OF $\hat{\underline{\beta}}$

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

• LINEAR TRANSFORMATIONS of GAUSSIAN RANDOM VECTORS

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

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

$$\mathbb{E}[\underline{T}] = \mathbb{E}[A\underline{Z} + \underline{b}] = A\,\mathbb{E}[\underline{Z}] + \underline{b} = A\underline{\mu} + \underline{b}$$
$$var(\underline{T}) = var(A\underline{Z} + b) = var(A\underline{Z}) = \mathbb{E}[(AZ - \mathbb{E}[AZ])(AZ - \mathbb{E}[AZ])^T]$$
$$= \mathbb{E}[(AZ - A\mu)(AZ - A\mu)^T] = \mathbb{E}[A(Z-\mu)(Z-\mu)^TA^T] = A\Sigma A^T$$

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

$$\mathbb{E}[\hat{\underline{\beta}}] = A\,\mathbb{E}[\underline{Y}] = (X^TX)^{-1}X^T\,\mathbb{E}[\underline{Y}] = \cancel{(X^TX)^{-1}}\,\cancel{X^TX}\underline{\beta} = \underline{\beta}$$
$$var(\hat{\underline{\beta}}) = var(A\underline{Y}) = A\,var(\underline{Y})\,A^T = A(\sigma^2 I_n)A^T = \sigma^2 AA^T = \sigma^2\,\cancel{(X^TX)^{-1}}\,\cancel{X^TX}\,((X^TX)^{-1})^T = \sigma^2(X^TX)^{-1}$$

$$\Rightarrow \quad \hat{\underline{\beta}} \sim N_p(\underline{\beta}, \sigma^2(X^TX)^{-1})$$

the marginal distribution of the j-th component is $\hat{\beta}_j \sim N_1(\beta_j\,;\ \sigma^2\,[(X^TX)^{-1}]_{(j,j)})$ $\quad j = 1,\dots,p$

the covariance between two components j and s $\quad cov(\hat{\beta}_j, \hat{\beta}_s) = \sigma^2[(X^TX)^{-1}]_{(j,s)}$ $\quad j,s = 1,\dots,p$

Notice that the variance is $\sigma^2(X^TX)^{-1}$: once again, we need $(\underline{x}_1, \dots, \underline{x}_p)$ to be linearly independent.

indeed $(X^TX)^{-1} = \frac{1}{det(X^TX)}\cdot[\dots]$

If they are linearly dependent, $det(X^TX) = 0$ and (X^TX) is not invertible

Remark: if they are "almost collinear" $(det(X^TX) \approx 0)$ the variance of the estimator explodes

$$var(\hat{\underline{\beta}}) = \sigma^2\cdot\underbrace{\frac{1}{det(X^TX)}}_{\to +\infty}\cdot[\dots]$$

• DISTRIBUTION OF THE RESIDUALS

$$\underline{e} = \underline{y} - \hat{\underline{y}} = \underline{y} - X\hat{\underline{\beta}} = \underline{y} - X\cdot(X^TX)^{-1}X^T\underline{y} = (I_n - X(X^TX)^{-1}X^T)\,\underline{y} = (I_n - P)\,\underline{y}$$

we called it P:
$P = X(X^TX)^{-1}X^T$ $\quad (n\times n)$ matrix

Let us study the corresponding random quantity $\underline{E}$

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

$$(I_n - P)X\underline{\beta} = X\underline{\beta} - PX\underline{\beta} = X\underline{\beta} - \underbrace{X\cancel{(X^TX)^{-1}}X^T}_{P}\cancel{X}\underline{\beta} = X\underline{\beta} - X\underline{\beta} = 0$$

Hence, $\underline{E} = (I_n - P)\underline{\varepsilon}$ $\qquad \underline{\varepsilon} \sim N_n(\underline{0}, \sigma^2 I_n)$

linear combination of a Gaussian is Gaussian $\to$ $\underline{E} \sim N_n(\dots,\dots)$

$\mathbb{E}[\underline{E}] = (I_n - P)\,\underline{0} = \underline{0}$

$var(\underline{E}) = (I_n - P)\,var(\underline{\varepsilon})\,(I_n - P)^T = \sigma^2(I_n - P)(I_n - P)^T = \sigma^2(I_n - P)$

$$\Rightarrow \quad \underline{E} \sim N_n(\underline{0}, (I_n - P)\sigma^2)$$

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

• DISTRIBUTION OF $\hat{\Sigma}^2$

Preliminary result

If $\underline{Z} \sim N_d(\underline{0}, \sigma^2 I_n)$; B a $(d\times d)$ matrix, symmetric, idempotent, with rank k $(1 \le k \le d)$

$$\Rightarrow \quad Q = \frac{1}{\sigma^2}\underline{Z}^TB\underline{Z} \qquad Q \sim \chi^2_k$$

$$\hat{\Sigma}^2 = \frac{1}{n}\underline{E}^T\underline{E}$$

$$\Rightarrow \quad n\hat{\Sigma}^2 = \underline{E}^T\underline{E} = ((I_n - P)\underline{\varepsilon})^T((I_n - P)\underline{\varepsilon})$$
$$= \underline{\varepsilon}^T(I_n - P)\underline{\varepsilon}$$

We have seen that $(I_n - P)$ is symmetric and idempotent.
Moreover, the rank of $(I_n - P)$ is equal to $(n-p)$.

$\Rightarrow$ applying the result we get

$$\frac{1}{\sigma^2}\cdot\underline{\varepsilon}^T(I_n - P)\underline{\varepsilon} \sim \chi^2_{n-p} \qquad \Rightarrow \qquad \frac{\underline{E}^T\underline{E}}{\sigma^2} = \frac{n\hat{\Sigma}^2}{\sigma^2} \sim \chi^2_{n-p}$$

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

Hence, we also find that $\mathbb{E}\left[\frac{n\hat{\Sigma}^2}{\sigma^2}\right] = n-p \quad \Rightarrow \quad \mathbb{E}[\hat{\Sigma}^2] = \frac{\sigma^2}{n}(n-p)$ $\quad$ biased

As usual, we obtain an unbiased estimator as

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

moreover, $\hat{\underline{\beta}} \perp\!\!\!\perp \hat{\Sigma}^2$ and $\hat{\underline{\beta}} \perp\!\!\!\perp S^2$