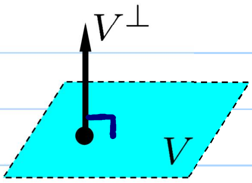
Normal Distribution and Sequential ANOVA ← Type I ANOVA

- Consider a linear model $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$, where
  - (may have more than one effects) $\boldsymbol{Y} \in \mathbb{R}^n$: an $n \times 1$ random vector and $\boldsymbol{Y} \sim \boldsymbol{N}(\boldsymbol{X\beta}, \sigma^2\boldsymbol{I})$ (↑ sample size)
  - $\boldsymbol{X} = [\boldsymbol{1} \;\; \boldsymbol{X}_1 \;\; \cdots \;\; \boldsymbol{X}_k]$ (intercept)
  - $\boldsymbol{\beta} = [\beta_0^T \;\; \beta_1^T \;\; \cdots \;\; \beta_k^T]^T$

  $\Rightarrow \boldsymbol{X\beta} = \underbrace{\boldsymbol{1}}_{\boldsymbol{X}_0}\beta_0 + \boldsymbol{X}_1\boldsymbol{\beta}_1 + \cdots + \boldsymbol{X}_k\boldsymbol{\beta}_k$, with $\boldsymbol{\mu}_0 = \boldsymbol{1}\beta_0$, $\boldsymbol{\mu}_1 = \boldsymbol{X}_1\boldsymbol{\beta}_1$, …, $\boldsymbol{\mu}_k = \boldsymbol{X}_k\boldsymbol{\beta}_k$ (nx1 vectors); $\mu_i \in \text{span}\{X_i\}$

  - (sequential ANOVA) $\boldsymbol{\varepsilon} \sim \boldsymbol{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$ ← $\varepsilon_1, \ldots, \varepsilon_n \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)$
- Define → $H_0^{(i)}: \beta_i = 0$ (V_{i-1}) vs. $H_1^{(i)}: \beta_i \neq 0$ (V_i), $i = 1, \ldots, k$.
  - V_i = the vector space generated by the column vectors of

    $A_i = [A_{i-1} \; X_i]$ $\quad \boldsymbol{A}_i \equiv [\boldsymbol{X}_0 \;\; \boldsymbol{X}_1 \;\; \cdots \;\; \boldsymbol{X}_i]$, $i = 0, 1, \ldots, k$.



    - V_i is called the *model space* of $\boldsymbol{A}_i$, and denoted by span$\{\boldsymbol{A}_i\}$
    - $V_0 \subset V_1 \subset \cdots \subset V_k = \text{span}\{\boldsymbol{X}\} \subset \mathbb{R}^n$
  - $V^\perp$ = orthogonal complement of the vector space V, i.e., → $\mathbb{R}^n = V \oplus V^\perp$

    (a vector space) $V^\perp = \{\boldsymbol{v} \in \mathbb{R}^n : \boldsymbol{v} \text{ is orthogonal to all the vectors in } V\}$

(check graph in LNp.2-25)

- $W_0 = \text{span}\{\boldsymbol{X}_0\} = V_0$, and for $i = 1, \ldots, k$,

  W_i: 隨機 ($H_0^{(i)}$) or 規律 ($H_1^{(i)}$)? for testing $H_0^{(i)}: \beta_i = 0$ in type I ANOVA.

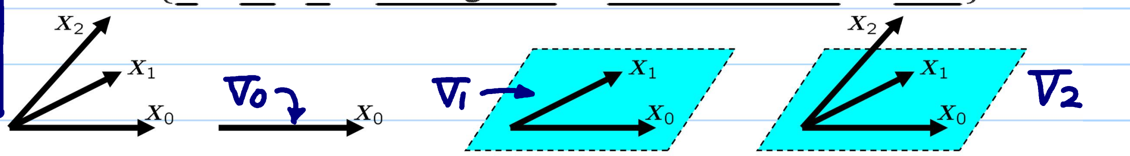
  W_i = orthogonal complement of V_{i-1} relative to V_i (**note**: $V_{i-1} \subset V_i$), i.e.,

  (a vector space) $W_i \;=\; V_i \ominus V_{i-1} \;=\; V_i \cap V_{i-1}^\perp$ ⟷ cf. $V_i \cap V_{i-1}^c$

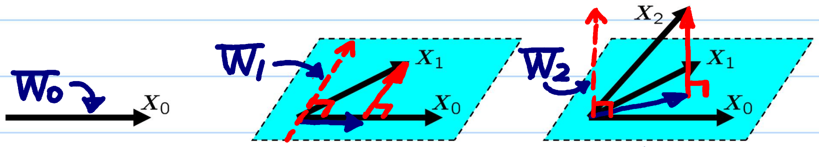
  $\quad = \{\boldsymbol{v} \in V_i : \boldsymbol{v} \text{ is orthogonal to all the vectors in } V_{i-1}\}$

  $V_2 \ominus \text{span}\{X_0, X_2\} \neq W_1$ in general (Type III ANOVA)

  span{A_i} ⊖ span{A_{i-1}} = span{[A_{i-1} X_i]} ⊖ span{A_{i-1}} ≠ span{X_i} in general

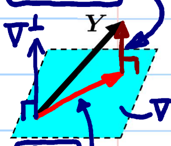


$W_1 \neq \text{span}\{X_1\}$, $W_2 \neq \text{span}\{X_2\}$



  - (direct sum) $V_i = W_0 \oplus W_1 \oplus \cdots \oplus W_i$ ⇒ $v \in V_i$ is uniquely represented $= \omega_0 + \omega_1 + \cdots + \omega_i$ ($\omega_0 \in W_0$, $\omega_1 \in W_1$, …, $\omega_i \in W_i$) (V_{i-1})
  - $\mathbb{R}^n = W_0 \oplus W_1 \oplus \cdots \oplus W_k \oplus V_k^\perp$ (**note**: $V_k^\perp$ is the residual space) (V_k) (mutually orthogonal)
  - $W_0 \perp W_1 \perp \cdots \perp W_k \perp V_k^\perp$ ⇒ orthogonality (↑ under the linear model in LNp.31.)
  - **Note**. In general, $W_i \neq \text{span}\{\boldsymbol{X}_i\}$. However, if $\boldsymbol{X}_0, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_k$ are mutually orthogonal, then $W_i = \text{span}\{\boldsymbol{X}_i\}$.
  - (usually equals the # of effects in X_i) $r_i \equiv \dim(W_i) = \dim(V_i) - \dim(V_{i-1})$.
  - Let $r = \sum_{i=0}^k r_i = \sum_{i=0}^k \dim(W_i) = \dim(V_k)$. Then, $\dim(V_k^\perp) = n - r$. (規律 … 隨機)

- Orthogonal projection of $\boldsymbol{Y}$ onto V_i's and W_i's
  - For a vector space $V \subset \mathbb{R}^n$, denote the *orthogonal projection matrix* of $\boldsymbol{Y}$ onto V by $\boldsymbol{P}_V$. Then, the orthogonal projection of $\boldsymbol{Y}$ onto V is $\boldsymbol{P}_V\boldsymbol{Y}$.

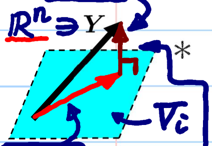


    * if $V = \text{span}\{\boldsymbol{A}\}$, then $\boldsymbol{P}_V = \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$
    * the orthogonal projection matrix onto $V^\perp$, denoted by $\boldsymbol{P}_{V^\perp}$, is $\boldsymbol{P}_{V^\perp} = \boldsymbol{I} - \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T = \boldsymbol{I} - \boldsymbol{P}_V$
  - Some properties of orthogonal projection matrix
    * A square matrix $\boldsymbol{P}$ is a *projection matrix* iff $\boldsymbol{P}^2 = \boldsymbol{P}$ (idempotent)
      · Idempotence implies $\boldsymbol{P}$ is a generalized inverse of $\boldsymbol{P}$ since

$$\boldsymbol{P}\boldsymbol{P}\boldsymbol{P} = \boldsymbol{P}^3 = \boldsymbol{P}.$$

    * A projection matrix $\boldsymbol{P}$ is *orthogonal* iff $\boldsymbol{P}^T = \boldsymbol{P}$ (symmetric)
    * If $\boldsymbol{P}$ is an orthogonal projection matrix onto V, then
      · $\boldsymbol{P}$ has $\dim(V)$ eigenvalues equal to 1 and the rest 0
      · $\boldsymbol{P}$ is diagonalizable, and there exists an orthogonal matrix $\boldsymbol{U}$ ($\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}$) such that $\boldsymbol{U}^T\boldsymbol{P}\boldsymbol{U} = \boldsymbol{\Lambda}$ is a diagonal matrix. (**Note.** Thus, $\boldsymbol{P} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$) Actually,
        ◇ the columns of $\boldsymbol{U}$ are orthonormal eigenvectors of $\boldsymbol{P}$, and
        ◇ the diagonal entries of $\boldsymbol{\Lambda}$ are the eigenvalues of $\boldsymbol{P}$

  - Since $V_i = \text{span}\{\boldsymbol{A}_i\}$, $\boldsymbol{P}_{V_i} = \boldsymbol{A}_i(\boldsymbol{A}_i^T\boldsymbol{A}_i)^{-1}\boldsymbol{A}_i^T$ and $\boldsymbol{P}_{V_i}\boldsymbol{Y} = \boldsymbol{A}_i(\boldsymbol{A}_i^T\boldsymbol{A}_i)^{-1}\boldsymbol{A}_i^T\boldsymbol{Y}$



    * For the model space V_i,

$$RSS_{V_i} = \|\boldsymbol{Y}\|^2 - \|\boldsymbol{P}_{V_i}\boldsymbol{Y}\|^2 = \boldsymbol{Y}^T\boldsymbol{Y} - (\boldsymbol{P}_{V_i}\boldsymbol{Y})^T\boldsymbol{P}_{V_i}\boldsymbol{Y} = \boldsymbol{Y}^T\boldsymbol{Y} - \boldsymbol{Y}^T\boldsymbol{P}_{V_i}^T\boldsymbol{P}_{V_i}\boldsymbol{Y}$$

$$= \boldsymbol{Y}^T\boldsymbol{Y} - \boldsymbol{Y}^T\boldsymbol{P}_{V_i}\boldsymbol{Y} = \boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{P}_{V_i})\boldsymbol{Y} = \|(\boldsymbol{I} - \boldsymbol{P}_{V_i})\boldsymbol{Y}\|^2$$

    * Since $W_i = V_i \cap V_{i-1}^\perp$ and $V_{i-1} \subset V_i$,

$$\boldsymbol{P}_{W_i} = (\boldsymbol{I} - \boldsymbol{P}_{V_{i-1}})\boldsymbol{P}_{V_i} = \boldsymbol{P}_{V_i} - \boldsymbol{P}_{V_{i-1}}\boldsymbol{P}_{V_i} = \boldsymbol{P}_{V_i} - \boldsymbol{P}_{V_{i-1}},$$

and $\boldsymbol{P}_{W_i}\boldsymbol{Y} = (\boldsymbol{I} - \boldsymbol{P}_{V_{i-1}})\boldsymbol{P}_{V_i}\boldsymbol{Y} = (\boldsymbol{P}_{V_i} - \boldsymbol{P}_{V_{i-1}})\boldsymbol{Y} = \boldsymbol{P}_{V_i}\boldsymbol{Y} - \boldsymbol{P}_{V_{i-1}}\boldsymbol{Y}$.

    * Since $\mathbb{R}^n = W_0 \oplus W_1 \oplus \cdots \oplus W_k \oplus V_k^\perp$ and $W_0 \perp W_1 \perp \cdots \perp W_k \perp V_k^\perp$,
      ◦ $\boldsymbol{Y} = \boldsymbol{P}_{W_0}\boldsymbol{Y} + \boldsymbol{P}_{W_1}\boldsymbol{Y} + \cdots + \boldsymbol{P}_{W_k}\boldsymbol{Y} + \boldsymbol{P}_{V_k^\perp}\boldsymbol{Y}$
      ◦ $\boldsymbol{P}_{W_0}\boldsymbol{Y} \perp \boldsymbol{P}_{W_1}\boldsymbol{Y} \perp \cdots \perp \boldsymbol{P}_{W_k}\boldsymbol{Y} \perp \boldsymbol{P}_{V_k^\perp}\boldsymbol{Y}$
      ◦ $\|\boldsymbol{Y}\|^2 = \|\boldsymbol{P}_{W_0}\boldsymbol{Y}\|^2 + \|\boldsymbol{P}_{W_1}\boldsymbol{Y}\|^2 + \cdots + \|\boldsymbol{P}_{W_k}\boldsymbol{Y}\|^2 + \|\boldsymbol{P}_{V_k^\perp}\boldsymbol{Y}\|^2$
    * When $\boldsymbol{X}_0, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_k$ are mutually orthogonal,

$$\boldsymbol{P}_{W_i} = \boldsymbol{P}_{V_i} - \boldsymbol{P}_{V_{i-1}} = \boldsymbol{X}_i(\boldsymbol{X}_i^T\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i^T.$$