Normal Distribution and Sequential ANOVA

Type I ANOVA

- Consider a linear model $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$, where
  - $\boldsymbol{Y} \in \mathbb{R}^n$: an $n \times 1$ random vector and $\boldsymbol{Y} \sim N(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{I})$

    sample size

    may have more than one effects

    $n \times 1$ vectors

  - $\boldsymbol{X} = [\boldsymbol{1} \quad \boldsymbol{X}_1 \quad \cdots \quad \boldsymbol{X}_k]$

    intercept

  - $\boldsymbol{\beta} = [\boldsymbol{\beta}_0^T \quad \boldsymbol{\beta}_1^T \quad \cdots \quad \boldsymbol{\beta}_k^T]^T$

    $\Rightarrow \boldsymbol{X\beta} = \overbrace{\underbrace{\boldsymbol{1}}_{\boldsymbol{X}_0} \boldsymbol{\beta}_0}^{\boldsymbol{\mu}_0} + \overbrace{\boldsymbol{X}_1 \boldsymbol{\beta}_1}^{\boldsymbol{\mu}_1} + \cdots + \overbrace{\boldsymbol{X}_k \boldsymbol{\beta}_k}^{\boldsymbol{\mu}_k}$

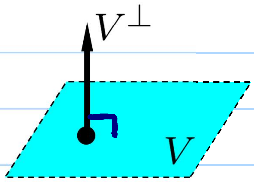
    $\boxed{\mu_i \in \text{span}\{X_i\}}$

    sequential ANOVA

  - $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$ ← $\varepsilon_1, \dots, \varepsilon_n \overset{iid}{\sim} N(0, \sigma^2)$

- Define → $H_0^{(i)}: \beta_i = 0 \ (V_{i-1})$ vs. $H_1^{(i)}: \beta_i \neq 0 \ (V_i)$, $i = 1, \dots, k$.
  - V_i = the vector space generated by the column vectors of

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



    * 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., → $\boxed{\mathbb{R}^n = V \oplus V^\perp}$

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

check graph in LNp.2-25

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

  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

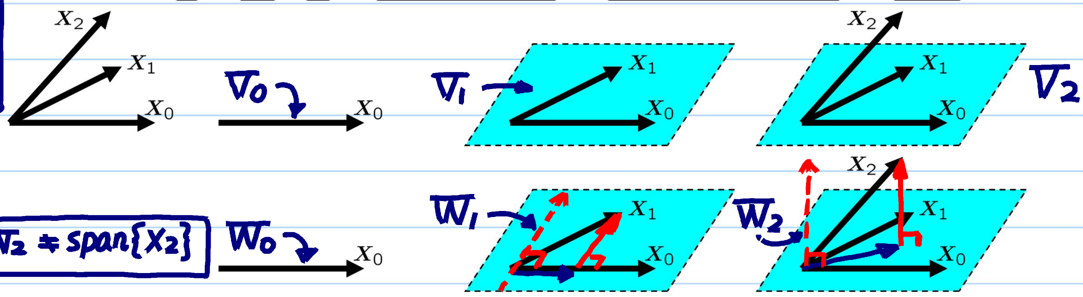
  $$W_i = V_i \ominus V_{i-1} = V_i \cap V_{i-1}^\perp \overset{\text{cf.}}{\longleftrightarrow} V_i \cap V_{i-1}^c$$

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

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

  span$\{A_i\}$ $\ominus$ span$\{A_{i-1}\}$ = span$\{[A_{i-1} \ X_i]\}$ $\ominus$ span$\{A_{i-1}\}$ $\neq$ 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

  * $W_0 \perp W_1 \perp \cdots \perp W_k \perp V_k^\perp$ ⇒ orthogonality ← under the linear model in LNp.31
  * **Note**. In general, $W_i \neq \text{span}\{\boldsymbol{X}_i\}$. However, if $\boldsymbol{X}_0, \boldsymbol{X}_1, \dots, \boldsymbol{X}_k$ are mutually orthogonal, then $W_i = \text{span}\{\boldsymbol{X}_i\}$.

  usually equals the # of effects in X_i

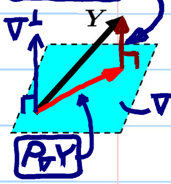
- $r_i \equiv \dim(W_i) = \dim(V_i) - \dim(V_{i-1})$.

  規律 隨機

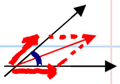
  * Let $r = \sum_{i=0}^k r_i = \sum_{i=0}^k \dim(W_i) = \dim(V_k)$. Then, $\dim(V_k^\perp) = n - r$.

- Orthogonal projection of $\boldsymbol{Y}$ onto V_i's and W_i's ← a linear transformation of $\boldsymbol{Y}$

  $P_{V^\perp}Y = (I-P_V)Y$

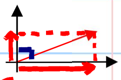
  - For a vector space $V \subset \mathbb{R}^n$, denote the *orthogonal projection matrix* of $\boldsymbol{Y}$ onto V by $\boldsymbol{P}_V$. Then, the orthogonal projection of $\boldsymbol{Y}$ onto V is $\boldsymbol{P}_V\boldsymbol{Y}$. ← an $n\times n$ matrix
    - ∗ if $V = \text{span}\{\boldsymbol{A}\}$, then $\boldsymbol{P}_V = \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$ ← Recall. Hat matrix in linear model
    - ∗ the orthogonal projection matrix onto $V^\perp$, denoted by $\boldsymbol{P}_{V^\perp}$, is $\boldsymbol{P}_{V^\perp} = \boldsymbol{I} - \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T = \boldsymbol{I} - \boldsymbol{P}_V$ → $P_{V^\perp}Y = (I-P_V)Y = Y - P_VY$

  - Some properties of orthogonal projection matrix
    - ⊛ A square matrix $\boldsymbol{P}$ is a *projection matrix* iff $\boldsymbol{P}^2 = \boldsymbol{P}$ (idempotent) ← iff
      - · Idempotence implies $\boldsymbol{P}$ is a generalized inverse of $\boldsymbol{P}$ since ← usually, a singular matrix

        $$\boldsymbol{P}\underbrace{\boldsymbol{P}}_{P^-}\boldsymbol{P} = \boldsymbol{P}^3 = \boldsymbol{P}.$$

        not necessarily orthogonal

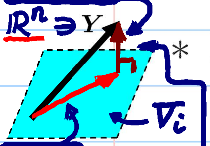
    - ⊛ A projection matrix $\boldsymbol{P}$ is *orthogonal* iff $\boldsymbol{P}^T = \boldsymbol{P}$ (symmetric)

      P is an orthogonal projection matrix ⟺ (iff)

    - ∗ If $\boldsymbol{P}$ is an orthogonal projection matrix onto V, then ← $n - \dim(V)$
      - ⊙ $\boldsymbol{P}$ has $\dim(V)$ eigenvalues equal to 1 and the rest 0 ← $\text{rank}(P) = \dim(V)$
      - · $\boldsymbol{P}$ 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. → $\Lambda = \begin{bmatrix}\lambda_1 & & 0\\ & \ddots & \\ 0 & & \lambda_n\end{bmatrix}$, $\lambda_i \in \{0, 1\}$

        (**Note.** Thus, $\boldsymbol{P} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$) Actually, ← eigen-decomposition
        - ◇ the columns of $\boldsymbol{U}$ are orthonormal eigenvectors of $\boldsymbol{P}$, and
        - ◇ the diagonal entries of $\boldsymbol{\Lambda}$ are the eigenvalues of $\boldsymbol{P}$

$P_{V^\perp}Y = (I-P_V)Y$



$\mathbb{R}^n \ni Y$

- 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 $\boldsymbol{P}_{V_i}\boldsymbol{Y} = \boldsymbol{A}_i(\boldsymbol{A}_i^T\boldsymbol{A}_i)^{-1}\boldsymbol{A}_i^T\boldsymbol{Y}$ ← LNp.31
  - ∗ For the model space V_i, ← i.e. consider the model: $Y \sim X_0\beta_0 + X_1\beta_1 + \cdots + X_i\beta_i$

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

    $= P_{V_i}P_{V_i} = P_{V_i}^2 = P_{V_i}$

    $P_{V_i^\perp}^T P_{V_i^\perp} = P_{V_i^\perp} =$

    $\mathbb{R}^n = V_i \oplus V_i^\perp$, $V_i = V_{i-1} \oplus W_i$

  - ⊙ Since $W_i = V_i \cap V_{i-1}^\perp$ and $V_{i-1} \subset V_i$, (exercise, prove that this is an orthogonal projection matrix. Hint. $P_{V_{i-1}}P_{V_i} = P_{V_i}P_{V_{i-1}} = P_{V_{i-1}}$)

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

    and $\boldsymbol{P}_{W_i}\boldsymbol{Y} = (\boldsymbol{I} - \boldsymbol{P}_{V_{i-1}})\boldsymbol{P}_{V_i}\boldsymbol{Y} = (\boldsymbol{P}_{V_i} - \boldsymbol{P}_{V_{i-1}})\boldsymbol{Y} = \boldsymbol{P}_{V_i}\boldsymbol{Y} - \boldsymbol{P}_{V_{i-1}}\boldsymbol{Y}$. ← check the graphs in LNp.32

    ✱ For a vector space V, $Y^TP_VY = \|P_VY\|^2$ is a quadratic form.

  - ∗ Since $\mathbb{R}^n = W_0 \oplus W_1 \oplus \cdots \oplus W_k \oplus V_k^\perp$ and $W_0 \perp W_1 \perp \cdots \perp W_k \perp V_k^\perp$,
    - ⊙ $\boldsymbol{Y} = \boldsymbol{P}_{W_0}\boldsymbol{Y} + \boldsymbol{P}_{W_1}\boldsymbol{Y} + \cdots + \boldsymbol{P}_{W_k}\boldsymbol{Y} + \boldsymbol{P}_{V_k^\perp}\boldsymbol{Y}$
    - ⊙ $\boldsymbol{P}_{W_0}\boldsymbol{Y} \perp \boldsymbol{P}_{W_1}\boldsymbol{Y} \perp \cdots \perp \boldsymbol{P}_{W_k}\boldsymbol{Y} \perp \boldsymbol{P}_{V_k^\perp}\boldsymbol{Y}$ ← cf. Type III ANOVA; check ✱ in LNp.32
    - ⊙ $\|\boldsymbol{Y}\|^2 = \|\boldsymbol{P}_{W_0}\boldsymbol{Y}\|^2 + \|\boldsymbol{P}_{W_1}\boldsymbol{Y}\|^2 + \cdots + \|\boldsymbol{P}_{W_k}\boldsymbol{Y}\|^2 + \|\boldsymbol{P}_{V_k^\perp}\boldsymbol{Y}\|^2$ ← cf.

    Partition of sums of squares in Type I ANOVA

  - ∗ When $\boldsymbol{X}_0, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_k$ are mutually orthogonal, ← $W_i = \text{span}\{X_i\}$

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

    $A_i = [A_{i-1}\ X_i]$

- Consider the sequential ANOVA: for $i=1,\ldots,k$, **Recall ANOVA in LNp.26**

Under the linear model in LNp.31 → $H_0^{(i)}: \boldsymbol{\beta}_i = \mathbf{0}\ (\omega_i = V_{i-1})$ vs. $H_A^{(i)}: \boldsymbol{\beta}_i \neq \mathbf{0}\ (\Omega_i = V_i)$ **Type III ANOVA**

(P1) $RSS_{\omega_i} - RSS_{\Omega_i} = RSS_{V_{i-1}} - RSS_{V_i} = \boldsymbol{Y}^T(\boldsymbol{I}-\boldsymbol{P}_{V_{i-1}})\boldsymbol{Y} - \boldsymbol{Y}^T(\boldsymbol{I}-\boldsymbol{P}_{V_i})\boldsymbol{Y}$

Sum of squares → check LNp.25, 26

$$= \boldsymbol{Y}^T(\boldsymbol{P}_{V_i}-\boldsymbol{P}_{V_{i-1}})\boldsymbol{Y} = \boldsymbol{Y}^T\boldsymbol{P}_{W_i}\boldsymbol{Y} = \boldsymbol{Y}^T\boldsymbol{P}_{W_i}^T\boldsymbol{P}_{W_i}\boldsymbol{Y} = \|\boldsymbol{P}_{W_i}\boldsymbol{Y}\|^2$$ cf.

and $df_{\omega_i} - df_{\Omega_i} = \dim(W_i) = r_i$. $P_{W_i}u_i$ may be $\approx \underline{0}$ even if $\underline{u_i}$ is not (e.g., $cor(u_i, u_{i-1}) \approx 1$)

may not be $\underline{0}$ even when $H_0^{(i)}: \underline{\beta_i} = \underline{0}\ (\Rightarrow \underline{u_i} = \underline{0})$ is true (e.g. $cor(u_i, u_k) \approx 1$)

(P2) $\boldsymbol{P}_{W_i}\boldsymbol{Y} = \boldsymbol{P}_{W_i}\boldsymbol{X\beta} + \boldsymbol{P}_{W_i}\boldsymbol{\varepsilon}$, where $\boxed{\underline{Y} = \underline{X}\underline{\beta} + \underline{\varepsilon}}$

– $\boldsymbol{P}_{W_i}\boldsymbol{X\beta} = (\boldsymbol{P}_{V_i}-\boldsymbol{P}_{V_{i-1}})(\boldsymbol{X}_0\boldsymbol{\beta}_0 + \cdots + \boldsymbol{X}_{i-1}\boldsymbol{\beta}_{i-1} + \boldsymbol{X}_i\boldsymbol{\beta}_i + \cdots + \boldsymbol{X}_k\boldsymbol{\beta}_k)$

$\because V_{i-1} = span\{x_0, \ldots, x_{i-1}\}$ and $W_i \subset V_{i-1}^{\perp}$, $\therefore P_{W_i}X_t = \underline{0}$ for $t = 0, 1, \ldots, i-1$

$$= (\boldsymbol{P}_{V_i}-\boldsymbol{P}_{V_{i-1}})(\boldsymbol{X}_i\boldsymbol{\beta}_i + \cdots + \boldsymbol{X}_k\boldsymbol{\beta}_k) = \boldsymbol{P}_{W_i}(\boldsymbol{u}_i + \cdots + \boldsymbol{u}_k)$$

* If $\boldsymbol{X}_0, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_k$ are mutually orthogonal, $\Rightarrow u_i \in span\{x_i\} = W_i$ cf.

$$\boldsymbol{P}_{W_i}\boldsymbol{X\beta} = \boldsymbol{P}_{W_i}\boldsymbol{u}_i = (\boldsymbol{X}_i(\boldsymbol{X}_i^T\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i^T)\boldsymbol{X}_i\boldsymbol{\beta}_i = \boldsymbol{X}_i\boldsymbol{\beta}_i = \boldsymbol{u}_i.$$

By (N1) in LNp.30 – $\boldsymbol{P}_{W_i}\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\boldsymbol{P}_{W_i}\boldsymbol{I}\boldsymbol{P}_{W_i}^T)$, where $\boldsymbol{P}_{W_i}\boldsymbol{I}\boldsymbol{P}_{W_i}^T = \boldsymbol{P}_{W_i}$ and $\boldsymbol{P}_{W_i} = \boldsymbol{P}_{W_i}^-$. $rank(P_{W_i}) = \dim(W_i) = r_i$

$\boxed{\varepsilon \sim N(\underline{0}, \sigma^2 I)}$ – Thus, $\boldsymbol{P}_{W_i}\boldsymbol{Y} \sim N(\boldsymbol{P}_{W_i}(\boldsymbol{u}_i + \cdots + \boldsymbol{u}_k), \sigma^2\boldsymbol{P}_{W_i})$ $(\sigma^2 P_{W_i})(\frac{1}{\sigma^2}P_{W_i})(\sigma^2 P_{W_i}) = \sigma^2 P_{W_i}$, $(\sigma^2 P_{W_i})^-$

(P3) $\|\boldsymbol{P}_{W_i}\boldsymbol{Y}\|^2 = \boldsymbol{Y}^T\boldsymbol{P}_{W_i}\boldsymbol{Y} = (\boldsymbol{X\beta}+\boldsymbol{\varepsilon})^T\boldsymbol{P}_{W_i}(\boldsymbol{X\beta}+\boldsymbol{\varepsilon})$

random variable
$$= (\boldsymbol{X\beta})^T\boldsymbol{P}_{W_i}(\boldsymbol{X\beta}) + 2(\boldsymbol{X\beta})^T\boldsymbol{P}_{W_i}\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}^T\boldsymbol{P}_{W_i}\boldsymbol{\varepsilon}$$

function of parameters $\underline{\beta}$
$$= \|\boldsymbol{P}_{W_i}\boldsymbol{X\beta}\|^2 + 2(\boldsymbol{X\beta})^T\boldsymbol{P}_{W_i}\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}^T\boldsymbol{P}_{W_i}\boldsymbol{\varepsilon},$$

FYI, $\|P_{W_i}Y\|^2/\sigma^2 \sim$ noncentral $\chi^2_{r_i}(\delta/\sigma^2)$ By (N6) in LNp.30

where $\boldsymbol{\varepsilon}^T\boldsymbol{P}_{W_i}\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^T\boldsymbol{P}_{W_i}^T\boldsymbol{P}_{W_i}\boldsymbol{P}_{W_i}\boldsymbol{\varepsilon} = \sigma^2(\boldsymbol{P}_{W_i}\boldsymbol{\varepsilon})^T(\boldsymbol{P}_{W_i}/\sigma^2)(\boldsymbol{P}_{W_i}\boldsymbol{\varepsilon}) \sim \sigma^2\chi^2_{r_i}$

(P4) $E(RSS_{V_{i-1}} - RSS_{V_i}) = E(\boldsymbol{Y}^T\boldsymbol{P}_{W_i}\boldsymbol{Y}) = E(\|\boldsymbol{P}_{W_i}\boldsymbol{Y}\|^2)$

E(sum of squares)
$$= \|\boldsymbol{P}_{W_i}\boldsymbol{X\beta}\|^2 + 2(\boldsymbol{X\beta})^T\boldsymbol{P}_{W_i}E(\boldsymbol{\varepsilon}) + E(\boldsymbol{\varepsilon}^T\boldsymbol{P}_{W_i}\boldsymbol{\varepsilon})$$

What if orthogonality exists? 未知
$$= \|\boldsymbol{P}_{W_i}(\boldsymbol{u}_i + \cdots + \boldsymbol{u}_k)\|^2 + r_i\sigma^2$$ 已知 $= \underline{0}$, $\sigma^2\chi^2_{r_i}$

– **Note**. If $\boldsymbol{M}$: a symmetric matrix, and $\boldsymbol{Z} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then

Z: a random variable $E(Z^2) = [E(Z)]^2 + Var(Z)$ cf.
$$E(\boldsymbol{Z}^T\boldsymbol{MZ}) = \boldsymbol{\mu}^T\boldsymbol{M\mu} + \text{trace}(\boldsymbol{M\Sigma}).$$

$Z^TMZ = (Z-\mu)^TM(Z-\mu) + 2\mu^TMZ - \mu^TM\mu$

$E(Z^TMZ) = 2\mu^TM\mu - \mu^TM\mu + E[(Z-\mu)^TM(Z-\mu)]$

$= E\{tr[(Z-\mu)^TM(Z-\mu)]\}$
$= E\{tr[M(Z-\mu)(Z-\mu)^T]\}$
$= tr\{M\, E[(Z-\mu)(Z-\mu)^T]\}$, $= \Sigma$

(P5) For the residual space $V_k^{\perp}$,

– $RSS_{V_k} = \boldsymbol{Y}^T(\boldsymbol{I}-\boldsymbol{P}_{V_k})\boldsymbol{Y} = \boldsymbol{Y}^T\boldsymbol{P}_{V_k^{\perp}}\boldsymbol{Y} = \|\boldsymbol{P}_{V_k^{\perp}}\boldsymbol{Y}\|^2$

$P_{W_i}Y$ cf. – $\boldsymbol{P}_{V_k^{\perp}}\boldsymbol{Y} = \underbrace{\boldsymbol{P}_{V_k^{\perp}}\boldsymbol{X\beta}}_{=\underline{0}} + \boldsymbol{P}_{V_k^{\perp}}\boldsymbol{\varepsilon} = \boldsymbol{P}_{V_k^{\perp}}\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\boldsymbol{P}_{V_k^{\perp}})$

$\|P_{W_i}Y\|^2$ cf. – $\|\boldsymbol{P}_{V_k^{\perp}}\boldsymbol{Y}\|^2 = \boldsymbol{\varepsilon}^T\boldsymbol{P}_{V_k^{\perp}}\boldsymbol{\varepsilon} \sim \sigma^2\chi^2_{n-r}$ $\dim(V_k^{\perp})$

$E(RSS_{V_{i-1}} - RSS_{V_i})$ cf. – $E(RSS_{V_k}) = E(\|\boldsymbol{P}_{V_k^{\perp}}\boldsymbol{Y}\|^2) = (n-r)\sigma^2$ 已知

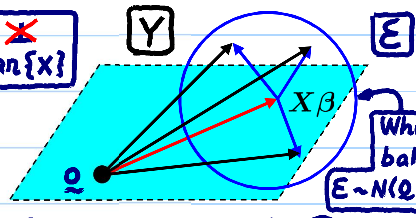
ε ~~$\in$~~ $span\{x\}$

Type III ANOVA cf.

(P6) Since $W_0 \perp W_1 \perp \cdots \perp W_k \perp V_k^{\perp}$,

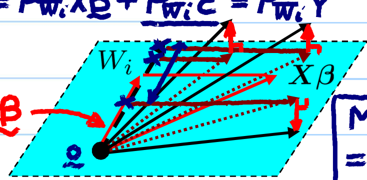
By (N4) & (N5) in LNp.30 $\Rightarrow \boldsymbol{P}_{W_0}\boldsymbol{Y}, \ldots, \boldsymbol{P}_{W_k}\boldsymbol{Y}, \boldsymbol{P}_{V_k^{\perp}}\boldsymbol{Y}$ are *independent* random vectors

$P_{W_i}P_{W_j}^T = \underline{0}$ for $i \neq j$ $\Rightarrow \|\boldsymbol{P}_{W_0}\boldsymbol{Y}\|^2, \ldots, \|\boldsymbol{P}_{W_k}\boldsymbol{Y}\|^2, \|\boldsymbol{P}_{V_k^{\perp}}\boldsymbol{Y}\|^2$ are *independent* random variables



Why a ball? $\varepsilon \sim N(\underline{0}, \sigma^2 I)$

$P_{W_i}Y = P_{W_i}X\beta + P_{W_i}\varepsilon = P_{W_i}\hat{Y}$



$P_{W_i}X\beta$ $\hat{Y}$ MSE $=$ bias^2 $+$ var

❖ **Further reading**: Seber and Lee (2003), *Linear Regression Analysis*, 2nd edition, Chapter 2.