(Bailey)

Section 2.4. orthogonal projection.

Q: What is orthogonal projection? (See the plot in LN. 2-6)

Q: Why is orthogonal projection important?

- many procedure of estimation & testing do nothing more than the decomposition of data vector into orthogonal pieces.
- orthogonality often corresponds to uncorrelated (become "independent" if more assumption, such as normality, is added)

Def: If W is a subspace of V, then the <u>orthogonal complement of W</u> is

$\{v \in V \mid v \text{ is orthogonal to all vectors in } W\}$

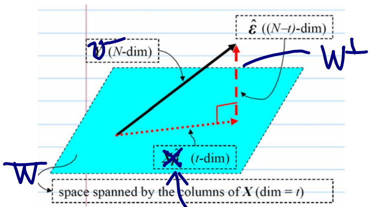
denoted by $W^{\perp}$

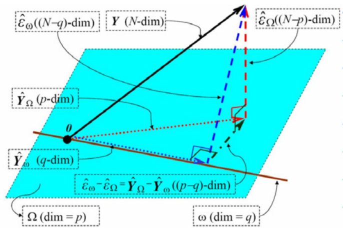
Theorem 2.3 *Let W be a subspace of V. Then the following hold.*

(i) $W^{\perp}$ *is also a subspace of* V.

(ii) $(W^{\perp})^{\perp} = W$.

(iii) $\dim(W^{\perp}) = \dim V - \dim W$.





(iv) V *is the* internal direct sum $W \oplus W^{\perp}$; *this means that given any vector* $\mathbf{v}$ *in* V *there is a unique vector* $\mathbf{x}$ *in* W *and a unique vector* $\mathbf{z}$ *in* $W^{\perp}$ *such that* $\mathbf{v} = \mathbf{x} + \mathbf{z}$. *We call* $\mathbf{x}$ *the* orthogonal projection *of* $\mathbf{v}$ *onto* W, *and write* $\mathbf{x} = P_W \mathbf{v}$. *See Figure 2.2*

(v) $P_{W^{\perp}}\mathbf{v} = \mathbf{z} = \mathbf{v} - \mathbf{x} = \mathbf{v} - P_W\mathbf{v}. = (I - P_W)v$

(vi) *For a fixed vector* $\mathbf{v}$ *in* V *and vector* $\mathbf{w}$ *in* W, $\sum_{\omega \in \Omega}(v_{\omega} - w_{\omega})^2 = \|\mathbf{v} - \mathbf{w}\|^2$. *As* $\mathbf{w}$ *varies over* W, *this sum of squares of differences is minimized when* $\mathbf{w} = P_W\mathbf{v}$. $\Rightarrow$ least square estimator

$\underset{w \in W}{\text{minimize}} \|v - w\|^2$

(vii) *If* $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ *is an orthogonal basis for* W *then*

different orthogonal effects.

$$P_W\mathbf{v} = \underbrace{\left(\frac{\mathbf{v}\cdot\mathbf{u}_1}{\underbrace{\mathbf{u}_1\cdot\mathbf{u}_1}_{=\|u_1\|^2}}\right)\mathbf{u}_1}_{\left(v \cdot \frac{u_1}{\|u_1\|}\right)\cdot\frac{u_1}{\|u_1\|}} + \left(\frac{\mathbf{v}\cdot\mathbf{u}_2}{\mathbf{u}_2\cdot\mathbf{u}_2}\right)\mathbf{u}_2 + \cdots + \left(\frac{\mathbf{v}\cdot\mathbf{u}_n}{\mathbf{u}_n\cdot\mathbf{u}_n}\right)\mathbf{u}_n.$$

(Bailey) Section 2.5. Linear model.

A linear model for unstructured plots

$$\underset{\text{from treatment}}{\mathbb{Y} = \underbrace{\tau}} + \underset{\text{from plot structure}}{\underbrace{\mathbb{Z}}}$$

Q: how to describe the distribution of $\mathbb{Z}$ for unstructured plots?

(i) $E(\mathbb{Z}) = \underset{\sim}{0}$ (Q: What if $E(\mathbb{Z}) = \mu \cdot \underbrace{\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}}_{= \mathbb{1}}$ & $\mu \neq 0$?)

(ii) $\mathrm{Var}(Z_\omega) = \sigma^2$, for $\omega \in \Omega$

(iii) $\mathrm{cov}(Z_\alpha, Z_\beta) = 0$, for different plots $\alpha, \beta \in \Omega$

(iv) $\tau \in V_T$ (but unknown)

$\Rightarrow E(\mathbb{Y}) = \underbrace{\tau}_{= X\beta}$, $\mathrm{cov}(\mathbb{Y}) = \sigma^2 I$.

$\mathbb{Y} = (\mu\mathbb{1} + \tau) + \underbrace{(\mathbb{Z} - \mu\mathbb{1})}_{\equiv \mathbb{Z}'}$

$E(\mathbb{Z}') = 0 \Leftarrow \mathbb{Z}'$

- If normality assumption is added, $\mathbb{Y} \sim N(\tau, \sigma^2 I) \Rightarrow$ independence

2-10

Theorem 2.4 *Assume that* $\mathbb{E}(\mathbf{Y}) = \tau$ *and that* $\mathrm{Cov}(\mathbf{Y}) = \sigma^2\mathbf{I}$. *Let W be a d-dimensional subspace of V. Then*

$E[Y'AY] = \mathrm{tr}[A\Sigma] + \theta' A \theta$, where $\theta = E[Y]$ & $\Sigma = \mathrm{cov}(Y)$.

(i) $\mathbb{E}(P_W\mathbf{Y}) = P_W(\mathbb{E}(\mathbf{Y})) = P_W\tau$;

(ii) $\mathbb{E}(\underbrace{\|P_W\mathbf{Y}\|^2}_{= Y'P_W'P_WY}) = \|P_W\tau\|^2 + d\sigma^2$. $\to$ related to "sum of square"

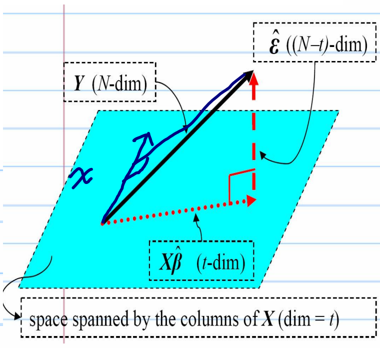
Theorem 2.5 *Assume that* $\mathbb{E}(\mathbf{Y}) = \tau \in V_T$ *and that* $\mathrm{Cov}(\mathbf{Y}) = \sigma^2\mathbf{I}$. *Let* $\mathbf{x}$ *and* $\mathbf{z}$ *be any vectors in* V_T. *Then*

(i) the $\overbrace{\textit{best (that is, minimum variance) linear unbiased estimator}}^{= \text{BLUE}}$ *of the scalar* $\mathbf{x}\cdot\tau$ *is* $\mathbf{x}\cdot\mathbf{Y}$; $\leftarrow$ Gauss-Markov Thm.

(ii) the variance of the estimator $\mathbf{x}\cdot\mathbf{Y}$ *is* $\|\mathbf{x}\|^2\sigma^2$; $\to \mathrm{Var}(x^T Y Y^T x) = x^T \sigma^2 I x = \sigma^2 x^T x$

(iii) the covariance of $\mathbf{x}\cdot\mathbf{Y}$ *and* $\mathbf{z}\cdot\mathbf{Y}$ *is* $(\mathbf{x}\cdot\mathbf{z})\sigma^2$.

$$x = \begin{bmatrix} 1/r_1 \\ \vdots \\ 1/r_1 \\ -1/r_2 \\ \vdots \\ -1/r_2 \\ 0 \end{bmatrix} \quad \tau = \begin{bmatrix} \tau_A \\ \vdots \\ \tau_A \\ \tau_B \\ \vdots \\ \tau_B \\ \vdots \end{bmatrix} \Rightarrow x \cdot \tau = \tau_A - \tau_B$$



2-11

(Bailey) Section 2.5. Estimation.

- Some properties:
  - Let $u_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ → corresponds to the r_i plots with treatment i $\Rightarrow$ (i) $u_i \cdot u_i = r_i$ (ii) for $v \in V$, $u_i \cdot v =$ sum of the values of v on those plots with treatment i.
  - Write
    (i) $u_i \cdot Y \equiv SUM_{T=i}$
    (ii) $u_i \cdot Y / r_i \equiv MEAN_{T=i}$
  - Let $u_0 = \underline{1}$, then (i) $u_0 = \sum_{i=1}^{t} u_i$
    (ii) $u_0 \cdot u_0 = N$
    (iii) For $v \in V$, $u_0 \cdot v = \sum_{\omega \in \Omega} v_\omega \Rightarrow \bar{v} \equiv u_0 \cdot v / N$
  - Write
    (i) $SUM \equiv u_0 \cdot Y \equiv N \cdot \bar{Y}$

- To estimate τ_i: Let $x_i = \frac{1}{r_i} u_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1/r_i \\ 1/r_i \\ \vdots \\ 1/r_i \\ 0 \\ \vdots \\ 0 \end{bmatrix} \Rightarrow x_i \cdot \tau = \tau_i$

  $\Rightarrow x_i \cdot Y = \frac{1}{r_i}(u_i \cdot Y) = MEAN_{T=i}$ → BLUE of τ_i

- To estimate $\sum_{i=1}^{t} \lambda_i \tau_i$: Let $x = \sum_{i=1}^{t} \frac{\lambda_i}{r_i} u_i = \begin{bmatrix} \lambda_1/r_1 \\ \lambda_1/r_1 \\ \lambda_1/r_1 \\ \lambda_2/r_2 \\ \lambda_2/r_2 \\ \vdots \end{bmatrix}$

  $\Rightarrow x \cdot \tau = \sum_{i=1}^{t} \frac{\lambda_i}{r_i} (u_i \cdot \tau) = \sum_{i=1}^{t} \frac{\lambda_i}{r_i} (r_i \cdot \tau_i)$

  $\Rightarrow$ BLUE of $x \cdot \tau$ is

  $x \cdot Y = \sum_{i=1}^{t} \lambda_i \cdot (MEAN_{T=i})$

- To estimate $\bar{\tau} = \sum_{i=1}^{t} \frac{r_i}{N} \cdot \tau_i \Rightarrow \lambda_i = \frac{r_i}{N}$

  $x \cdot Y = \sum_{i=1}^{t} \frac{r_i}{N} \cdot (MEAN_{T=i}) = \frac{SUM}{N} = \bar{Y}$

- Now we look at Theorem 2.4 with $W = V_T$. Since $\tau \in V_T$, we have

$$P_{V_T}\tau = \tau = \sum_{i=1}^{t} \tau_i \mathbf{u}_i.$$

$\{u_1, \cdots, u_t\}$ is an orthogonal basis.

$\hat{Y}$

Proposition 2.1 and Theorem 2.3(vii) show that

$\hat{\tau}_i$ ||

$$P_{V_T}\mathbf{Y} = \sum_{i=1}^{t} \left(\frac{\mathbf{Y} \cdot \mathbf{u}_i}{\mathbf{u}_i \cdot \mathbf{u}_i} \right) \mathbf{u}_i = \sum_{i=1}^{t} \frac{\text{SUM}_{T=i}}{r_i} \mathbf{u}_i = \sum_{i=1}^{t} \text{MEAN}_{T=i} \mathbf{u}_i.$$

Theorem 2.4(i) confirms that this is an unbiased estimator of τ.

(Bailey) Section 2.7. Sum of Square

Def: Let W be a subspace of V

(i) For $\mathbf{Y} \in V$, the sum of square for W means $\|P_W \mathbf{Y}\|^2$

(ii) the degree of freedom for W is $\dim(W)$

(iii) the mean square for W is $\|P_W \mathbf{Y}\|^2 / \dim(W)$

(iv) the expected mean square for W is

$$EMS(W) = E(\|P_W \mathbf{Y}\|^2 / \dim(W))$$

- First we apply these ideas with $W = V_T$. Since

$$P_{V_T}\mathbf{Y} = \sum_{i=1}^{t} \frac{\text{SUM}_{T=i}}{r_i}\mathbf{u}_i,$$

the sum of squares for V_T is equal to

$$(P_{V_T}Y)' = \left(\sum_{i=1}^{t} \frac{\text{SUM}_{T=i}}{r_i}\mathbf{u}_i\right) \cdot \left(\sum_{j=1}^{t} \frac{\text{SUM}_{T=j}}{r_j}\mathbf{u}_j\right) = (P_{V_t}Y)$$

$\to \sum_{i=1}^{t}\sum_{j=1}^{t} \hat{\tau}_i u_i \cdot \hat{\tau}_j \cdot u_j$

Now, $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ whenever $i \neq j$, so this sum of squares

$\boxed{\mathbb{1} \in V_T}$

$$= \sum_{i=1}^{t} \frac{\text{SUM}^2_{T=i}}{r_i^2} \underbrace{\mathbf{u}_i \cdot \mathbf{u}_i}_{=r_i} = \sum_{i=1}^{t} \frac{\text{SUM}^2_{T=i}}{r_i}.$$

The quantity $\sum_i(\text{sum}^2_{T=i}/r_i)$ is called the *crude sum of squares for treatments*, which may be abbreviated to CSS(treatments).

The number of degrees of freedom for V_T is simply the dimension of V_T, which is equal to t.

The mean square for V_T is equal to

$$\sum_{i=1}^{t} \frac{\text{SUM}^2_{T=i}}{r_i} \Bigg/ t.$$

Theorem 2.4(ii) shows that

$\tau \in V_T$

$$\mathbb{E}(\|P_{V_T}\mathbf{Y}\|^2) = \|\underbrace{P_{V_T}\tau}_{=\tau}\|^2 + t\sigma^2 = \sum_{i=1}^{t} r_i\tau_i^2 + t\sigma^2,$$

because $P_{V_T}\tau = \sum_{i=1}^{t}\tau_i\mathbf{u}_i$. Hence the expected mean square for V_T is equal to $\sum r_i\tau_i^2/t + \sigma^2$.

$\hookrightarrow E(\|P_{V_T}Y\|^2/t)$

- Secondly we apply the ideas with $W = V_T^{\perp}$. By Theorem 2.3(v),

$$
\begin{aligned}
P_W\mathbf{y} &= \mathbf{y} - P_{V_T}\mathbf{y} \\
&= \mathbf{y} - \sum_{i=1}^{t} \hat{\tau}_i \mathbf{u}_i \rightarrow \hat{y} \\
&= \text{data vector} - \text{vector of fitted values} \\
&= \text{residual vector},
\end{aligned}
$$

$P_W = (I - P_{V_T})$

- $\|P_W Y\|^2$ = sum of squares of the residuals

$$\underbrace{\sum_{\omega\in\Omega} y_\omega^2}_{\sum_{\omega\in\Omega}(y_\omega-\bar{y})^2} = \|\mathbf{y}\|^2 = \|P_{V_T}\mathbf{y}\|^2 + \|P_W\mathbf{y}\|^2.$$

↳ total sum of square

- $\dim(V_T^{\perp}) = \dim V - \dim V_T = \boxed{N-t}$

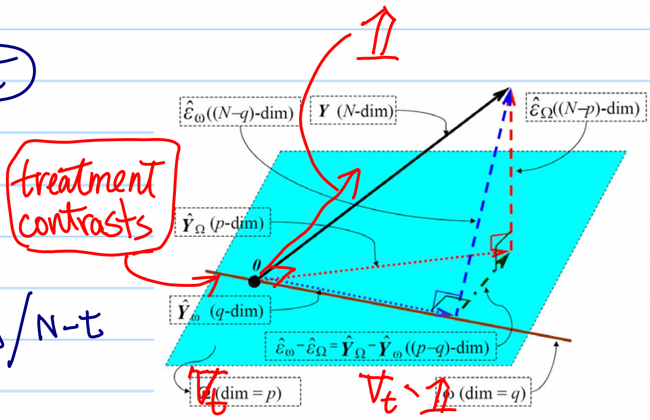
degree of freedom of residuals

- mean square

an unbiased estimator of σ^2 ← $MS(\text{residual}) = \|P_W Y\|^2 / N-t$

- $EMS(\text{residual}) = \{\underbrace{\|P_W \tau\|^2}_{0} + (N-t)\sigma^2\} / N-t$

$= \sigma^2$



(Bailey) Section 2.8. Variance

- To estimate $\sum_{i=1}^{t} \lambda_i \tau_i$, $x = \sum_{i=1}^{t} \frac{\lambda_i}{r_i} u_i$

BLUE: $x \cdot Y = \sum_{i=1}^{t} \lambda_i \, MEAN_{T=i}$

$$
\begin{aligned}
Var(x\cdot Y) = \|\mathbf{x}\|^2\sigma^2 &= \left(\sum_{i=1}^{t}\frac{\lambda_i}{r_i}\mathbf{u}_i\right)\cdot\left(\sum_{j=1}^{t}\frac{\lambda_j}{r_j}\mathbf{u}_j\right)\sigma^2 \\
&= \sum_{i=1}^{t}\frac{\lambda_i^2}{r_i^2}(\mathbf{u}_i\cdot\mathbf{u}_i)\sigma^2 \\
&= \left(\sum_{i=1}^{t}\frac{\lambda_i^2}{r_i}\right)\sigma^2.
\end{aligned}
$$

For different λ_i's, the "best" choice of r_i's is different.
⇒ choice of r_i's is corresponding to the choice of designs, i.e $T: \Omega \to \mathcal{T}$

- Two cases:
  - To estimate τ_{i_0} ⇒ $\lambda_{i_0} = 1$ and $\lambda_j = 0$ for $j \neq i_0$
    ⇒ $\left(\sum_{i=1}^{t} \lambda_i^2 / r_i\right)\sigma^2 = \sigma^2 / r_{i_0}$
  - To estimate $\tau_{i_0} - \tau_{i_1}$ ⇒ $\lambda_{i_0} = 1$, $\lambda_{i_1} = -1$ and $\lambda_j = 0$ for $\begin{cases} j \neq i_0 \\ j \neq i_1 \end{cases}$
    ⇒ $\left(\sum_{i=1}^{t} \frac{\lambda_i^2}{r_i}\right)\sigma^2 = \sigma^2\left(\frac{1}{r_{i_0}} + \frac{1}{r_{i_1}}\right)$