$$\underset{n\times 1}{y} = \underset{n\times p}{X}\,\underset{p\times 1}{\beta} \qquad \hat{\beta} = \underset{?}{(X^T X)^*} X^T y = Cy$$

THE LINEAR MODEL

- ## Projections

Theorem: Let $\boldsymbol{a}: n \times 1$ be a non-zero vector and let $\boldsymbol{d}: n \times 1$ be any vector. Then $\boldsymbol{d}$ can be uniquely expressed as:

$$\boldsymbol{d} = \boldsymbol{b_1} + \boldsymbol{b_2}$$

where $\boldsymbol{b_1}$ depends on $\boldsymbol{a}$ and $\boldsymbol{b_2}$ is orthogonal to $\boldsymbol{a}$. Furthermore, $\boxed{\boldsymbol{b_1} = c\boldsymbol{a}}$ with $c = \frac{\boldsymbol{d'a}}{\boldsymbol{a'a}}$ and $\boldsymbol{b_2} = \boldsymbol{d} - \boldsymbol{b_1}$.

Proof:

$$a^T b_1 + \cancel{a^T b_2}^{\,0}$$

Let $\boldsymbol{d} = \boldsymbol{b_1} + \boldsymbol{b_2} = c\boldsymbol{a} + \boldsymbol{b_2}$ where $\boldsymbol{b_2} \perp \boldsymbol{a}$. Then $\boldsymbol{a'd} = c\boldsymbol{a'a} + \boldsymbol{0}$ or $c = \frac{\boldsymbol{d'a}}{\boldsymbol{a'a}}$ since $\boldsymbol{a'a} \neq \boldsymbol{0}$.

Conversely, if $c = \frac{\boldsymbol{d'a}}{\boldsymbol{a'a}}$ and $\boldsymbol{b_2} = \boldsymbol{d} - \boldsymbol{b_1}$, then $\boldsymbol{b_2} = \boldsymbol{d} - c\boldsymbol{a}$. Therefore

$\boldsymbol{b_2} \perp \boldsymbol{a}$ and since c is unique, $\boldsymbol{b_2}$ is also unique.

The vector $\boldsymbol{b_1}$ is the projection of $\boldsymbol{d}$ on $\boldsymbol{a}$ and $\boldsymbol{b_2}$ is the component of $\boldsymbol{d}$ orthogonal to $\boldsymbol{a}$. The square of the length of the projection of $\boldsymbol{d}$ on $\boldsymbol{a}$ is

$$\boldsymbol{b_1'b_1} = c^2\boldsymbol{a'a} = \frac{(\boldsymbol{d'a})^2}{\boldsymbol{a'a}}.$$

Theorem: Let V be a vector space and let $\boldsymbol{d}$ be any vector. Then $\boldsymbol{d}$ can be uniquely expressed as:

$$\boldsymbol{d} = \boldsymbol{c} + \boldsymbol{e}$$

where $\boldsymbol{c} \in V$ and $\boldsymbol{e} \perp V$. If V is generated by the columns of $\boldsymbol{A}: n \times m$, then

$$\boldsymbol{c} = \boldsymbol{A}(\boldsymbol{A'A})^*\boldsymbol{A'd} \quad \text{and} \quad \boldsymbol{e} = (\boldsymbol{I} - \boldsymbol{A}(\boldsymbol{A'A})^*\boldsymbol{A'})\boldsymbol{d}$$

Proof:

$$A^T d = A^T A c_1 + \cancel{A^T e}^{\,0}$$

Let $\boldsymbol{d} = \boldsymbol{c} + \boldsymbol{e}$ where $\boldsymbol{c} \in V$ and $\boldsymbol{e} \perp V$. Then $\boldsymbol{d} = \boldsymbol{Ac_1} + \boldsymbol{e}$ for some $\boldsymbol{c_1}$. In other words

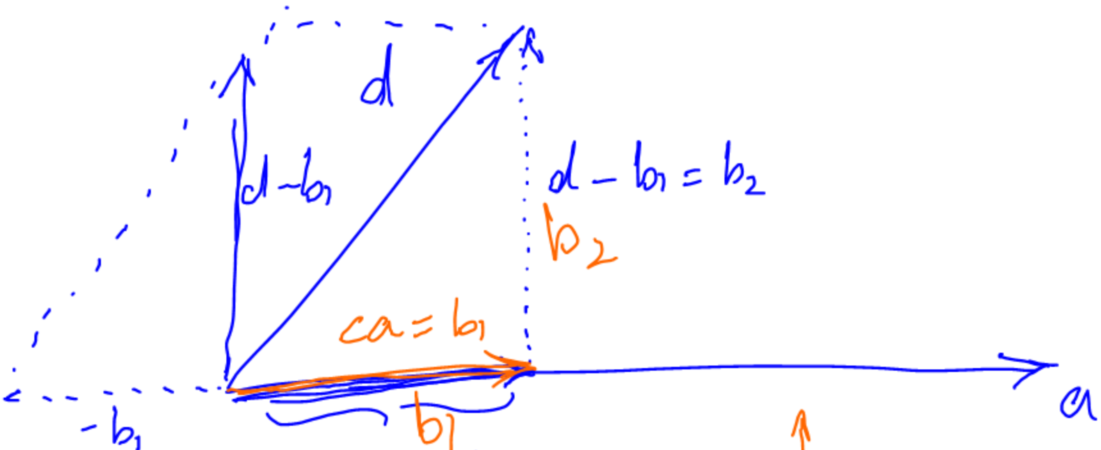
$$\boldsymbol{A'Ac_1} = \boldsymbol{A'd} \quad \text{since} \quad \boldsymbol{A'e} = \boldsymbol{0}.$$

These equations to determine $\boldsymbol{c_1}$ are consistent, because $\text{rank}\,(\boldsymbol{A'A}) = \text{rank}\,(\boldsymbol{A'A}, \boldsymbol{A'd})$. A solution of $\boldsymbol{A'Ac_1} = \boldsymbol{A'd}$ is

$$c_1 = (A^T A)^* A^T d \qquad d = Ac_1 = A(A^T A)^* A^T d$$

$d = b_1 + b_2$

$A = (a_1, a_2 - -, a_m)$



$b_1 = \text{Proj}_a d = ca$

$c =$

$$c_1 = (A'A)^* A'd$$

for any generalized inverse $(A'A)^*$ of $A'A$. Therefore

$$c = Ac_1 = A(A'A)^* A'd$$

and

$$e = d - c = (I - A(A'A)^* A')d.$$

Suppose now that a solution c_1 exists to

$$A'Ac_1 = A'd.$$

Let $c = Ac_1$ and $e = d - c$, then

$$d = c + e$$

where $c \in V$ and $e \perp V$, since $A^Te = A^Td - A^Tc \overset{c=Ac_1}{=\!=} A^Td - A^TAc_1 \overset{A^Td = A^TAc_1}{=\!=} A^TAc_1 - A^TAc_1 = 0$

$$A'e = A'd - A'c = A'd - A'Ac_1 = \mathbf{0}.$$

To prove the uniqueness of the expression, let

$$d = Ac_0 + e_0$$

with $e_0 \perp V$. Then

$$A'd = A'Ac_0 = A'Ac_1.$$

Therefore $A'A(c_0 - c_1) = \mathbf{0}$ and $(c_0 - c_1)'A'A(c_0 - c_1) = \mathbf{0}$. The last identity is the sum of squares of the elements of $A(c_0 - c_1)$, which is zero. In other words, $A(c_0 - c_1) = \mathbf{0}$ and $Ac_0 = Ac_1$. Furthermore,

$$e_0 = d - \underset{Ac_1}{Ac_0} = d - Ac_1 = e$$

and the expression is unique.

The vector c is the projection of d on V and e is the component of d which is orthogonal to V. Thus, $c = Ac_1$ is the projection of d ($A'Ac_1 = A'd$)

Proj $\hat{y}$ compare $\hat{y} = X\hat{\beta} = Py$

Conclusion: The equations

$y = x\beta$

$$A'Ax = A'y$$

$x^Tx\beta = x^Ty$

proj y

imply that Ax is the projection of y on the vector space generated by the columns of A. Thus, $Ax = A(A'A)^* A'y$ is the projection of y. And $A(A'A)^* A'$ is the projection matrix.

$P = A(A^TA)^* A^T$

Note: Let $V^{\perp}$ represent the vector space orthogonal to V. Since V is generated by the columns of $\boldsymbol{A}: n \times m$ the $\dim(V) + \dim(V^{\perp}) = n$.

Theorem: The projection matrix $\boldsymbol{P} = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{*}\boldsymbol{A}'$ is unique, symmetrical and idempotent. (Unique with respect to the choice of the generalized inverse and unique with respect to the specific $\boldsymbol{A}$. The only condition is that the vector space V must be generated by the columns of $\boldsymbol{A}$.)

Proof:

1. Uniqueness: Suppose that $\boldsymbol{P_1}$ is such a projection matrix as well. Then

$$c = \boldsymbol{P_1 d} = \boldsymbol{Pd} \quad \forall \boldsymbol{d} \quad \Longrightarrow \quad \boldsymbol{P_1} = \boldsymbol{P}$$

2. Symmetry: It follows from (1) and the fact that $(\boldsymbol{A}'\boldsymbol{A})^{*\prime}$ is also a generalized inverse of $\boldsymbol{A}'\boldsymbol{A}$. Therefore

$(AB)^T = B^T A^T$

$$P^T = P' = (\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{*}\boldsymbol{A}')' = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{*\prime}\boldsymbol{A}' = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{*}\boldsymbol{A}' = P$$

3. Idempotent: $\boldsymbol{P^2 d} = \boldsymbol{P}(\boldsymbol{Pd}) = \boldsymbol{Pd}$ since $\boldsymbol{Pd} \in V \quad \forall \boldsymbol{d}$. Therefore $\boldsymbol{P^2} = \boldsymbol{P}$.

P

$$P^2 = A(A^TA)^*A^TA(A^TA)^*A^T = A(A^TA)^*A^T = P$$

$(I-P)(I-P) = (I-P)$

Corollary:

1. If $\boldsymbol{A} = (\boldsymbol{A_1}, \boldsymbol{A_2})$ with $\boldsymbol{A_1'A_2} = \boldsymbol{0}$, it also follows that

$$\boldsymbol{P} = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{*}\boldsymbol{A}' = \boldsymbol{A_1}(\boldsymbol{A_1'A_1})^{*}\boldsymbol{A_1'} + \boldsymbol{A_2}(\boldsymbol{A_2'A_2})^{*}\boldsymbol{A_2'}.$$

2. If the columns of $\boldsymbol{A}$ are for instance mutually orthogonal, it follows that:

(a)

$a_i^T a_j = 0 \quad i \neq j$

$$\boldsymbol{P} = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{*}\boldsymbol{A}' = (\boldsymbol{a_1}, \boldsymbol{a_2}, \cdots, \boldsymbol{a_m}) \begin{pmatrix} (\boldsymbol{a_1'a_1})^{-1} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & (\boldsymbol{a_2'a_2})^{-1} & \cdots & \boldsymbol{0} \\ \cdots & \cdots & \cdots & \cdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & (\boldsymbol{a_m'a_m})^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{a_1'} \\ \boldsymbol{a_2'} \\ \cdots \\ \boldsymbol{a_m'} \end{pmatrix}$$

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}^T = (b_1^T \ b_2^T)$$

$$= \sum_{i=1}^{m} \boldsymbol{a_i}(\boldsymbol{a_i'a_i})^{-1}\boldsymbol{a_i'}.$$

(b) The projection of the vector $\boldsymbol{y}$ on V is:

nice representation

$$\hat{y} = \boldsymbol{Py} = \underbrace{\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{*}\boldsymbol{A}'\boldsymbol{y}}_{\text{proj } y} = \sum_{i=1}^{m} \boldsymbol{a_i}(\boldsymbol{a_i'a_i})^{-1}\boldsymbol{a_i'y} = \sum_{i=1}^{m} \frac{\boldsymbol{a_i'y}}{\boldsymbol{a_i'a_i}}\boldsymbol{a_i}$$

which is the sum of the projections on the individual mutually orthogonal vectors.

(c) The square of the length of the projection of $\boldsymbol{y}$ on V is:

$$\left(\text{Proj}_V y\right)^2 = \boldsymbol{y}'\boldsymbol{P}\boldsymbol{y} = \left|\sum_{i=1}^{m} \frac{(\boldsymbol{a}_i'\boldsymbol{y})^2}{\boldsymbol{a}_i'\boldsymbol{a}_i}.\right|$$

$$(Py)^T(Py) = y^T P^2 y = y^T P y$$

The projection matrix $\boldsymbol{P}$ can be calculated easily by using a convenient basis for V. The following theorem can for instance be used.

Theorem: Suppose that V with $\text{rank}\,(V) = r$ is generated by the columns of $\boldsymbol{A} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_r)$. It is always possible to select a mutually orthogonal basis for V, say the columns of $\boldsymbol{B} = (\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_r)$, in such a way that $\boldsymbol{b}_s$ only depends on the first s columns of $\boldsymbol{A}$.

Proof:

Assume the vector $\boldsymbol{a}_2$. It can be expressed uniquely as

$b_1 = c a_1$

$b_1 = 0 \Rightarrow \frac{a_2^T a_1}{a_1^T a_1} = 0 \Rightarrow a_2^T a_1 = 0 \Rightarrow a_1 \perp a_2 \; X.$

$$\boldsymbol{a}_2 = \boldsymbol{b}_1 + \boldsymbol{b}_2$$

where $\boldsymbol{b}_1$ is dependent on $\boldsymbol{a}_1$ and $\boldsymbol{b}_2$ is orthogonal to $\boldsymbol{a}_1$. The first two vectors $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ can be replaced by $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ in the basis, without changing V. If $\boldsymbol{b}_1 = \boldsymbol{0}$, let $\boldsymbol{b}_1 = \boldsymbol{a}_1$ and $\boldsymbol{b}_2 = \boldsymbol{a}_2$. The vector $\boldsymbol{a}_3$ can be uniquely expressed as

$\text{Proj}_{b_1} a_3 \quad \text{Proj}_{b_2} a_3$

$$\boldsymbol{a}_3 = \boldsymbol{b} + \boldsymbol{b}_3$$

where $\boldsymbol{b}$ depends on $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ and where $\boldsymbol{b}_3$ is orthogonal to $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$. It follows that

$b_1 = c a_1$

$$\boldsymbol{b} = \frac{\boldsymbol{a}_3'\boldsymbol{b}_1}{\boldsymbol{b}_1'\boldsymbol{b}_1}\boldsymbol{b}_1 + \frac{\boldsymbol{a}_3'\boldsymbol{b}_2}{\boldsymbol{b}_2'\boldsymbol{b}_2}\boldsymbol{b}_2$$

and $b_2 \perp a_1$

$$\boldsymbol{b}_3 = \boldsymbol{a}_3 - \boldsymbol{b}.$$

This process can be continued until all the $\boldsymbol{a}_i\boldsymbol{'s}$ are replaced by $\boldsymbol{b}_i\boldsymbol{'s}$ in the basis.

It follows that:

$$\boldsymbol{P} = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^*\boldsymbol{A}' = \boldsymbol{B}(\boldsymbol{B}'\boldsymbol{B})^*\boldsymbol{B}' = (\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_r)\begin{pmatrix} (\boldsymbol{b}_1'\boldsymbol{b}_1)^{-1} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & (\boldsymbol{b}_2'\boldsymbol{b}_2)^{-1} & \cdots & \boldsymbol{0} \\ \cdots & \cdots & \cdots & \cdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & (\boldsymbol{b}_r'\boldsymbol{b}_r)^{-1} \end{pmatrix}\begin{pmatrix} \boldsymbol{b}_1' \\ \boldsymbol{b}_2' \\ \cdots \\ \boldsymbol{b}_r' \end{pmatrix}$$

$$= \sum_{i=1}^{r} \boldsymbol{b}_i(\boldsymbol{b}_i'\boldsymbol{b}_i)^{-1}\boldsymbol{b}_i'.$$

Examples:

$a_1 \quad a_2$ $\quad a_1 = (2 \quad 1 \quad 3)^T \quad a_2 = (0 \quad 1 \quad 1)^T \quad a_1^T a_2 = 4 \neq 0$

1. Assume $\boldsymbol{A} = \begin{pmatrix} 2 & 0 \\ 1 & 1 \\ 3 & 1 \end{pmatrix}$. Then, it follows that:

$$b_1 = \frac{a_2^T a_1}{a_1^T a_1} a_1 = \frac{4}{14}\begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix} = \frac{2}{7}\begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}$$

$$\boldsymbol{b_1} = \frac{2}{7}\begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix} \text{ and } \boldsymbol{b_2} = \boldsymbol{a_2} - \boldsymbol{b_1} = \frac{1}{7}\begin{pmatrix} -4 \\ 5 \\ 1 \end{pmatrix}$$

$b_1^T b_2 = -8 + 5 + 3 = 0$

$\Rightarrow b_1 \perp b_2$

The projection matrix is: $P = \sum_{i=1}^{2} b_i (b_i^T b_i)^{-1} b_i^T$

$$\boldsymbol{P} = \frac{1}{14}\begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}(2 \quad 1 \quad 3) + \frac{1}{42}\begin{pmatrix} -4 \\ 5 \\ 1 \end{pmatrix}(-4 \quad 5 \quad 1)$$

$$= \frac{1}{14}\begin{pmatrix} 4 & 2 & 6 \\ 2 & 1 & 3 \\ 6 & 3 & 9 \end{pmatrix} + \frac{1}{42}\begin{pmatrix} 16 & -20 & -4 \\ -20 & 25 & 5 \\ -4 & 5 & 1 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

A better choice for $\boldsymbol{A}$ would be $\boldsymbol{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}$ since we can replace the first column with $\frac{1}{2}(\boldsymbol{a_1} - \boldsymbol{a_2})$

3×2

in $\boldsymbol{A} = \begin{pmatrix} 2 & 0 \\ 1 & 1 \\ 3 & 1 \end{pmatrix}$. Then, we have

$$\boldsymbol{b_1} = \frac{1}{2}\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

$$\boldsymbol{b_2} = \boldsymbol{a_2} - \boldsymbol{b_1} = \frac{1}{2}\begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}$$

The projection matrix is:

$$\boldsymbol{P} = \frac{1}{2}\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}(1 \quad 0 \quad 1) + \frac{1}{6}\begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}(-1 \quad 2 \quad 1)$$

$$= \frac{1}{2}\begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix} + \frac{1}{6}\begin{pmatrix} 1 & -2 & -1 \\ -2 & 4 & 2 \\ -1 & 2 & 1 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

An alternative calculation of $\boldsymbol{P}$ follows from the fact that

$A = \begin{pmatrix} 2 & 0 \\ 1 & 1 \\ 3 & 1 \end{pmatrix}$

$\text{rank}\,(V^{\perp}) = 1.\, V^{\perp}$ is for instance generated by

$\dim V + \dim V^{\perp} = 3$

$\dim V^{\perp} = 3 - 2 = 1$

$a_1^T e = 0 \;\&\; a_2^T e = 0$

$$\boldsymbol{e} = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$$

$e \in V^{\perp} \Rightarrow a_i^T e = 0 \quad i = 1, 2$

Therefore $\quad P(I - P) = P - P^2 = P - P = 0$

$$\boldsymbol{I} - \boldsymbol{P} = \boldsymbol{e}(\boldsymbol{e}'\boldsymbol{e})^{-1}\boldsymbol{e}'$$

$$= \frac{1}{3}\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}(1 \quad 1 \quad -1) = \frac{1}{3}\begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix}$$

and

$$\boldsymbol{P} = \frac{1}{3}\begin{pmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

2. For the previous Example 2 it follows that:

$$\boldsymbol{A} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 3 & 1 & 1 \end{pmatrix}$$

$\dim V + \dim V^{\perp} = 3$

An orthogonal basis for V is the columns of

indep

$\dim V^{\perp} = 3 - 2 = 1$

$$\boldsymbol{B} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}$$

$a_i^T e = 0$

The projection matrix is:

$$\boldsymbol{P} = \frac{1}{2}\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}(1 \quad 0 \quad 1) + \frac{1}{3}\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}(-1 \quad 1 \quad 1)$$

$$= \frac{1}{2}\begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix} + \frac{1}{3}\begin{pmatrix} 1 & -1 & -1 \\ -1 & 1 & 1 \\ -1 & 1 & 1 \end{pmatrix} = \frac{1}{6}\begin{pmatrix} 5 & -2 & 1 \\ -2 & 2 & 2 \\ 1 & 2 & 5 \end{pmatrix}$$

Again it follows that $rank\,(V^{\perp}) = 1$ and $V^{\perp}$ is generated by

$$\boldsymbol{e} = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$$

Therefore

$$\boldsymbol{I} - \boldsymbol{P} = \boldsymbol{e}(\boldsymbol{e}'\boldsymbol{e})^{-1}\boldsymbol{e}'$$

$$= \frac{1}{6}\begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}(1 \quad 2 \quad -1) = \frac{1}{6}\begin{pmatrix} 1 & 2 & -1 \\ 2 & 4 & -2 \\ -1 & -2 & 1 \end{pmatrix}$$

and

$$\boldsymbol{P} = \frac{1}{6}\begin{pmatrix} 5 & -2 & 1 \\ -2 & 2 & 2 \\ 1 & 2 & 5 \end{pmatrix}$$

$E y = \mu$

- ## Sums of Squares of Linear Sets

$\sum_{i=1}^{n} c_i y_i$

Let $\boldsymbol{y}: n \times 1 \sim (\boldsymbol{\mu}, \sigma^2\boldsymbol{I})$. The sum of squares of the linear function $\boldsymbol{c}'\boldsymbol{y}$ is defined as the square of the length of the projection of $\boldsymbol{y}$ on $\boldsymbol{c}$, namely:

$\mathrm{Proj}_c\, y$

$\mathrm{var}(y)$

$(c'y)' = y'c$

$$s^2 = \left(\frac{\boldsymbol{c}'\boldsymbol{y}}{\boldsymbol{c}'\boldsymbol{c}}\boldsymbol{c}\right)'\left(\frac{\boldsymbol{c}'\boldsymbol{y}}{\boldsymbol{c}'\boldsymbol{c}}\boldsymbol{c}\right)$$
$$= \frac{(\boldsymbol{c}'\boldsymbol{y})^2}{\boldsymbol{c}'\boldsymbol{c}}.$$

constant

$\mathrm{tr}\, V(y) = E y'y - (Ey)'(Ey)$

Note that this sum of squares is the same as the sum of squares of the linear function $d\boldsymbol{c}'\boldsymbol{y}$, for any non-zero constant d. This sum of squares is therefore only dependent on the vector space generated by **c**.

matrix

$\mathrm{var}(c^T y) = c^T \mathrm{Var}(y)\, c$
$= \sigma^2 c^T c$

It follows that:

$\mathrm{Var}(y) = E\, y y^T - (Ey)(Ey)^T$

$$E(s^2) = E(\boldsymbol{c}'\boldsymbol{y})^2/\boldsymbol{c}'\boldsymbol{c}$$
$$= \{\mathrm{var}\,(\boldsymbol{c}'\boldsymbol{y}) + [E(\boldsymbol{c}'\boldsymbol{y})]^2\}/\boldsymbol{c}'\boldsymbol{c}$$
$$= \sigma^2 + [\boldsymbol{c}'E(\boldsymbol{y})]^2/\boldsymbol{c}'\boldsymbol{c}$$

The expected value of s^2 is therefore $\boldsymbol{\sigma}^2$ plus a term which is obtained by replacing the random variables by their expected values in s^2. Note that:

$$E(s^2) = \sigma^2 \Leftrightarrow E(\boldsymbol{c}'\boldsymbol{y}) = 0$$

$n \times 1$

The sum of squares of a linear set is defined similarly. Suppose that $\boldsymbol{C}: n \times m = (\boldsymbol{c_1}, \boldsymbol{c_2}, \cdots, \boldsymbol{c_m})$ and $\text{rank}\ (\boldsymbol{C}) = r$. Let V_C be the vector space generated by the columns of $\boldsymbol{C}$. Let L_C be the linear set of all linear functions with coefficient vectors in V_C :

$$L_C = \{\boldsymbol{c}'\boldsymbol{y} : \boldsymbol{c} \in V_C\}.$$

L_C is generated by the linear functions $\boldsymbol{c}_1'\boldsymbol{y}, \cdots, \boldsymbol{c}_m'\boldsymbol{y}$ just as V_C is generated by $\boldsymbol{c_1}, \cdots, \boldsymbol{c_m}$. The linear functions are dependent or independent corresponding to dependence or independence of the coefficient vectors. Any r independent linear functions in L_C will therefore generate the linear set L_C. We say that L_C has r degrees of freedom.

The sum of squares of the linear set L_C is defined as the square of the length of the projection of $\boldsymbol{y}$ on V_C, namely:

$$y^T C(C^T C)^{*T} C^T C (C^T C)^* C^T y = y^T C (C^T C)^* C^T y$$

$$S^2 = \{\underbrace{\boldsymbol{C}(\boldsymbol{C}'\boldsymbol{C})^*\boldsymbol{C}'}_{P_C}\boldsymbol{y}\}'\{\boldsymbol{C}(\boldsymbol{C}'\boldsymbol{C})^*\boldsymbol{C}'\boldsymbol{y}\} = \boldsymbol{y}'\boldsymbol{C}(\boldsymbol{C}'\boldsymbol{C})^*\boldsymbol{C}'\boldsymbol{y}$$

since $\boldsymbol{C}(\boldsymbol{C}'\boldsymbol{C})^*\boldsymbol{C}'$ is unique, symmetrical and idempotent.

The expression for S^2 is independent of the specific functions $\boldsymbol{Cy}$ which generate L_C. Suppose that the columns of $\boldsymbol{B} = (\boldsymbol{b_1}, \boldsymbol{b_2}, \cdots, \boldsymbol{b_r})$ form a mutually orthogonal basis for V_C, so that $L_B = L_C$. It then follows that:

$$S^2 = \boldsymbol{y}'\boldsymbol{C}(\boldsymbol{C}'\boldsymbol{C})^*\boldsymbol{C}'\boldsymbol{y} = \boldsymbol{y}'\boxed{\boldsymbol{B}(\boldsymbol{B}'\boldsymbol{B})^*\boldsymbol{B}'}\boldsymbol{y}$$

$$= \boldsymbol{y}'(\boldsymbol{b_1}, \boldsymbol{b_2}, \cdots, \boldsymbol{b_r})\begin{pmatrix} (\boldsymbol{b}_1'\boldsymbol{b}_1)^{-1} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & (\boldsymbol{b}_2'\boldsymbol{b}_2)^{-1} & \cdots & \boldsymbol{0} \\ \cdots & \cdots & \cdots & \cdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & (\boldsymbol{b}_r'\boldsymbol{b}_r)^{-1} \end{pmatrix}\begin{pmatrix} \boldsymbol{b}_1' \\ \boldsymbol{b}_2' \\ \cdots \\ \boldsymbol{b}_r' \end{pmatrix}\boldsymbol{y}$$

$$= \sum_{i=1}^{r} \frac{(\boldsymbol{b}_i'\boldsymbol{y})^2}{\boldsymbol{b}_i'\boldsymbol{b}_i} = \sum_{i=1}^{r} s_i^2$$

where s_i^2 is the sum of squares of $\boldsymbol{b}_i'\boldsymbol{y}$. Also,

$$E(S^2) = \sum_{i=1}^{r} E\left(s_i^2\right) = r\sigma^2 + \underbrace{\sum_{i=1}^{r} (\boldsymbol{b}_i'E(\boldsymbol{y}))^2/\boldsymbol{b}_i'\boldsymbol{b}_i}_{S_\mu^2}.$$

It follows that:

$$E(S^2) = r\sigma^2 + S_\mu^2$$

where S_μ^2 is a term obtained by replacing the random variables in S^2 with their expected values.

$S_\mu^2 = 0$

Note that $E(S^2) = r\sigma^2 \Leftrightarrow$ if and only if $E(\boldsymbol{B}'\boldsymbol{y}) = \boldsymbol{0}$.

But $E(\boldsymbol{B}'\boldsymbol{y}) = \boldsymbol{0}$ if and only if $E(\boldsymbol{C}'\boldsymbol{y}) = \boldsymbol{0}$, since $\boldsymbol{B} = \boldsymbol{C}\boldsymbol{D}_1$ for some $\boldsymbol{D}_1$ and $\boldsymbol{C} = \boldsymbol{B}\boldsymbol{D}_2$ for some $\boldsymbol{D}_2$.

Suppose that V_{C_1} and V_{C_2} with ranks r_1 and r_2 respectively, are mutually orthogonal vector spaces with corresponding linear sets L_{C_1} and L_{C_2}. Therefore $\boldsymbol{C}_1'\boldsymbol{C}_2 = \boldsymbol{0}$. The linear sets L_{C_1} and L_{C_2} are orthogonal. Let

$$V_C = V_{C_1} \oplus V_{C_2} \text{ and } L_C = L_{C_1} \oplus L_{C_2}.$$

The implication is that V_C is generated by the columns of $\boldsymbol{C} = (\boldsymbol{C}_1, \boldsymbol{C}_2)$. If S_i^2 is the sum of squares of L_{C_i} and S^2 is the sum of squares of L_C, it then follows that:

$$S^2 = \boldsymbol{y}'\boldsymbol{C}(\boldsymbol{C}'\boldsymbol{C})^*\boldsymbol{C}'\boldsymbol{y} = \boldsymbol{y}'\boldsymbol{C}_1(\boldsymbol{C}_1'\boldsymbol{C}_1)^*\boldsymbol{C}_1'\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{C}_2(\boldsymbol{C}_2'\boldsymbol{C}_2)^*\boldsymbol{C}_2'\boldsymbol{y} = S_1^2 + S_2^2$$

The same line of reasoning can be used to show that if $V_{C_i}, i = 1, \cdots, k$ are mutually orthogonal vector spaces with corresponding linear sets $L_{C_i}, i = 1, \cdots, k$ and

$$V_C = \bigoplus V_{C_i} \text{ and } L_C = \bigoplus L_{C_i}$$

in other words V_C is generated by the columns of $\boldsymbol{C} = (\boldsymbol{C}_1, \boldsymbol{C}_2, \cdots, \boldsymbol{C}_k)$, then

$$S^2 = \boldsymbol{y}'\boldsymbol{C}(\boldsymbol{C}'\boldsymbol{C})^*\boldsymbol{C}'\boldsymbol{y} = \sum_{i=1}^{k} \boldsymbol{y}'\boldsymbol{C}_i(\boldsymbol{C}_i'\boldsymbol{C}_i)^*\boldsymbol{C}_i'\boldsymbol{y} = \sum_{i=1}^{k} S_i^2$$

with S_i^2 the sum of squares of L_{C_i}.

Consider the linear set

$$L_C = \{\boldsymbol{c}'\boldsymbol{y} \ : \ \boldsymbol{c} \in V_C\}.$$

with $\boldsymbol{C}: n \times m = (\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_m)$, rank $(\boldsymbol{C}) = r$ and V_C the vector space generated by the columns of $\boldsymbol{C}$. The mean square for L_C is

$ES^2 = r\sigma^2 + S_\mu^2$ $$S^2/r$$

with expected value

$$\sigma^2 + S_\mu^2/r.$$

Examples

1. Suppose that $\boldsymbol{y}: 4 \times 1 \sim (\boldsymbol{\mu}, \sigma^2 \boldsymbol{I})$ with

$$\boldsymbol{\mu} = \begin{pmatrix} \alpha \\ \alpha \\ \beta \\ \beta \end{pmatrix}$$

Let

orthogonal $\quad\quad \perp \quad\quad C_1^T C_2 = 0$

$$y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} \qquad \boldsymbol{C_1} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}, \ \boldsymbol{C_2} = \begin{pmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \end{pmatrix} \text{ and } \boldsymbol{C} = (\boldsymbol{C_1}, \boldsymbol{C_2}).$$

$c_1 \quad c_2$

The sums of squares of the linear sets L_{C_1} and L_{C_2} are

$$S_1^2 = y^T C_1 (C_1^T C_1)^* C_1 y = y^T c_1 (c_1^T c_1)^* c_1 y + y^T c_2 (c_2^T c_2)^* c_2 y = \frac{(c_1^T y)^2}{c_1^T c_1} + \frac{(c_2^T y)^2}{c_2^T c_2} =$$

$$S_1^2 = \frac{1}{2}(y_1 + y_2)^2 + \frac{1}{2}(y_3 + y_4)^2$$

and

$$y^T C_2 (C_2^T C_2)^* C_2 y$$

$$S_2^2 = \frac{1}{2}(y_1 - y_2)^2 + \frac{1}{2}(y_3 - y_4)^2.$$

respectively. The total sum of squares is

$$S_1^2 + S_2^2 = S^2 = y_1^2 + y_2^2 + y_3^2 + y_4^2 = S_1^2 + S_2^2.$$

The expected values are:

$E(S_1^2) = 2\sigma^2 + 2\alpha^2 + 2\beta^2$

$$E(S_2^2) = 2\sigma^2$$

$$\rightarrow E(S^2) = 4\sigma^2 + 2\alpha^2 + 2\beta^2.$$

2. Suppose that $\boldsymbol{y}: n \times 1 \sim (\boldsymbol{\mu}, \sigma^2 \boldsymbol{I})$. The set of all linear functions of $\boldsymbol{y}$ is generated by $\boldsymbol{Iy}$, with sum of squares:

$$S^2 = \boldsymbol{y}'\boldsymbol{I}(\boldsymbol{I}'\boldsymbol{I})^*\boldsymbol{I}'\boldsymbol{y} = \boldsymbol{y}'\boldsymbol{y} = \sum_{i=1}^{n} y_i^2.$$

Note also that the linear functions $y_1, y_2, \cdots y_n$ form a mutually orthogonal basis for this linear set. The sum of squares, S^2, is therefore the same as the total of the individual sums of squares of the $y_i{}'s$, and the sum of squares of

$$y_i = (\underbrace{0,0,\cdots,0}_{0},\overset{e_i}{1},\underbrace{0,\cdots,0}_{0})\boldsymbol{y}$$

is

$$s_i^2 = y_i^2.$$

Let L_E be the set of linear functions orthogonal to $\bar{y}$. The sum of squares of L_E is

$$S_E^2 = \boldsymbol{y}'\boldsymbol{y} - n\bar{y}^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2,$$

the mean square of L_E is

$$S_E^2/(n-1)$$

with expected value

$$\sigma^2 + \frac{1}{n-1}\sum_{i=1}^{n} (\mu_i - \bar{\mu})^2.$$

Also,

$$E(n\bar{y}^2) = \sigma^2 + n\bar{\mu}^2.$$

↓ Don't forget the last page

Furthermore: last example

Consider the set of all linear functions of y is generated by $I_n y \Rightarrow C = I_n$ & $\dim(V_C) = n$

Suppose $L_{C_1} = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n} 1_n^T y$ s.t. $1_n = (1, 1, \dots, 1)^T$, with one degree of freedom since V_{C_1} generated by vector 1_n.

$$S_1^2 = s_1^2 = \frac{(1_n^T y)^2}{1_n^T 1_n} = \frac{(n\bar{y})^2}{n} = n\bar{y}^2 \Rightarrow E(S_1^2) = E(n\bar{y}^2) = \sigma^2 + n\bar{\mu}^2 \qquad \bar{\mu} = \frac{1}{n} 1_n^T \mu$$

Since L_E is orthogonal to $\bar{y}$, L_{C_2} say with $n-1$ degrees of freedom

refers to error

I call

$V_C = V_{C_1} \oplus V_{C_2} \Rightarrow d.f.(L_{C_2}) = n-1$

$\Rightarrow V_{C_2}$ is generated by $n-1$ vector orthogonal to 1_n.

$$\Rightarrow S_2^2 = S_E^2 = S_C^2 - s_1^2 = y^T y - n\bar{y}^2$$

Indeed I used the following to derive above

Suppose that V_{C_1} and V_{C_2} with ranks r_1 and r_2 respectively, are mutually orthogonal vector spaces with corresponding linear sets L_{C_1} and L_{C_2}. Therefore $\boldsymbol{C}_1'\boldsymbol{C}_2 = \boldsymbol{0}$. The linear sets L_{C_1} and L_{C_2} are orthogonal. Let

$$V_C = V_{C_1} \oplus V_{C_2} \text{ and } L_C = L_{C_1} \oplus L_{C_2}.$$