The Geometric Interpretation of Missing Observations

BU-243-M | Robert Jacobsen | June, 1967

Abstract

This paper presents the geometric interpretation of missing observations in the general linear model.

Bicmetrics Unit, Department of Plant Breeding and Biometry, Cornell University, Ithaca, New York.

The Geometric Interpretation of Missing Observations

BU-243-M

Robert Jacobsen

June, 1967

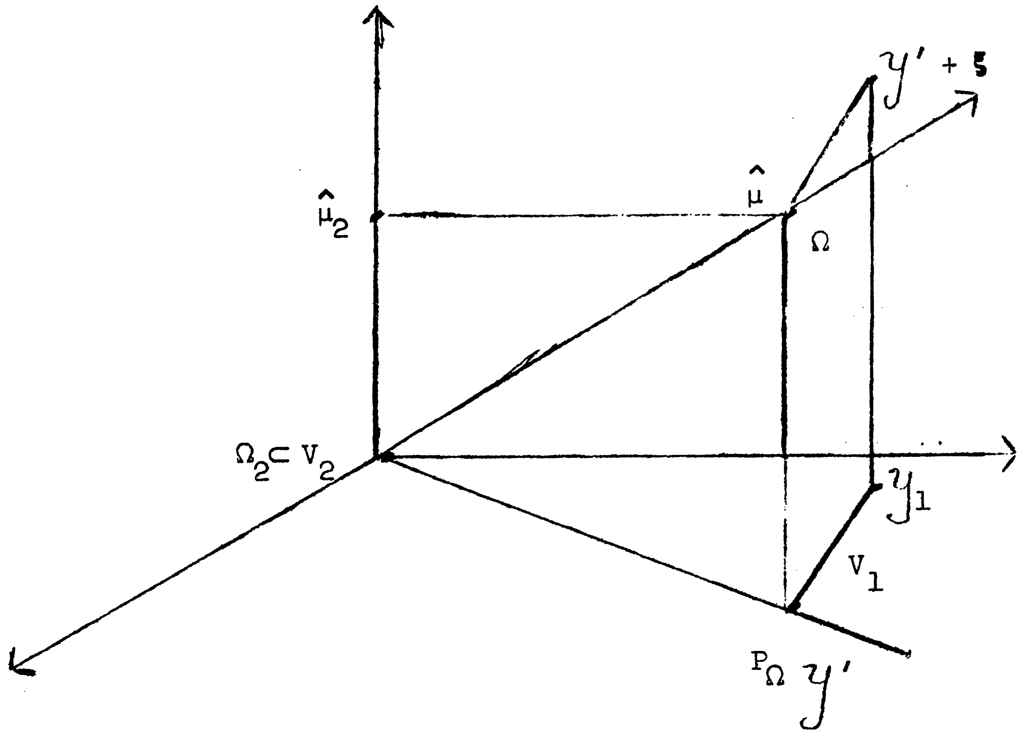
The observation space V is decomposed into $V_1 \oplus V_2$, $V_1 \perp V_2$. We only observe $\mathcal{Y}' \in V_1$. V_2 is the space of missing observations.

Let Ω be the original mean space, with $\Omega \not\subset V_1$, $\Omega \not\subset V_2$, $\Omega \neq V$. If $\mu \in \Omega$, $\mathcal{Y}'$ has expectation $P_{V_1} \mu = \mu_1$, which runs around in $\Omega_1 = P_{V_1} \Omega$, the mean space corresponding to those observations observed. Denote $P_{V_2} \Omega = \Omega_2$. We seek $P_{\Omega_1} \mathcal{Y}'$, the b.l.u.e. of $E\, \mathcal{Y}'$.

Now we could compute $P_{\Omega_1} \mathcal{Y}'$ directly, but the missing observation method utilizes the fact that P_Ω, P_{V_1}, and P_{V_2} are known explicitly.

We want an easy way of computing a point $\xi \in V_2$ (the missing observations) s.t. $P_{V_1} P_\Omega (\xi + \mathcal{Y}') = P_{\Omega_1} \mathcal{Y}'$, for all $\mathcal{Y}' \in V_1$. Thus, by completing the missing observations with ξ, applying the standard analysis and then projecting the least squares estimate onto V_1, we get the least squares estimate of $\mu_1 = E\, \mathcal{Y}'$, based on the observed $\mathcal{Y}'$.

Biometrics Unit, Department of Plant Breeding and Biometry, Cornell University, Ithaca, New York.



<u>Lemma</u>.

Let $L(y')$ be the space of solutions $z \in V_2$ of the equation

$$P_{V_2} P_\Omega (y' + z) = z, \text{ for each } y' \in V_1.$$

Let $L^*(y')$ be the space of solns. $z \in V_2$ to the equation

$$P_{V_1} P_\Omega (y' + z) = P_{\Omega_1} y'.$$

Then, for all $y' \in V_1$, $L(y') \neq \phi$, and $L(y') = L^*(y')$.

If $\dim \Omega_1 < \dim \Omega$, $L(y')$ has dim h, $h = \dim \Omega - \dim \Omega_1$.

Proof:

(i) $z, z' \in L(\mathcal{Y}') \Rightarrow z - z' = P_{V_2} P_{\Omega} (z - z')$.

By Bessel's inequality,

$\|z - z'\| \geq \|P_{V_2} P_{\Omega} (z - z')\|$, with equality holding iff $P_{V_2} P_{\Omega}$ operates as the identity map on $z - z'$. Hence, $z - z' \in \Omega \cap V_2$.

(ii) let $z, z' \in L^*(\mathcal{Y}')$. Then

$$0 = P_{V_1} P_{\Omega} (\mathcal{Y}' + z) - P_{V_1} P_{\Omega} (\mathcal{Y}' + z')$$
$$= P_{V_1} P_{\Omega} (z - z').$$ So $z - z' \in \text{Ker } P_{V_1} P_{\Omega}$.

But $\text{Ker } P_{V_1} = V_2$. So $\text{Ker } P_{V_1} P_{\Omega} = V_2 \cap \Omega$.

(iii) So $\dim L^*(\mathcal{Y}') = \dim L(\mathcal{Y}')$.

(iv) For each $\mu_1 \in \Omega_1$, $\mu_1 \neq 0$, let $S(\mu_1) =$

$P_{V_1}^{-1} \{\mu_1\} \cap \Omega$. $S(\mu_1) \neq \phi$, $\forall\ \mu_1 \neq 0$.

$\therefore$ For each $\mathcal{Y}' \in V_1$, $\exists\ \hat{\mu} \neq 0$ s.t. $\hat{\mu} \in S(P_{\Omega_1} \mathcal{Y}')$.
But $\hat{\mu} = P_{V_1} \hat{\mu} + P_{V_2} \hat{\mu} = P_{\Omega_1} \mathcal{Y}' + \hat{\mu}_2$, say.

Now $\hat{\mu} = P_{\Omega}\,(\hat{\mu}) = P_{\Omega}\,(P_{\Omega_1} y' + \hat{\mu}_2)$

$= P_{\Omega}\,(P_{\Omega_1} y' - y') + P_{\Omega}\,(y' + \hat{\mu}_2)$

$= P_{\Omega}\,(y' + \hat{\mu}_2)$, for $P_{\Omega_1} y' - y' + \Omega_1$ and Ω_2,

and $\Omega \subset \Omega_1 \oplus \Omega_2$, so $P_{\Omega_1} y' - y' + \Omega$.

Thus, $\hat{\mu}_2 = P_{V_2}\,\hat{\mu} = P_{V_2}\,P_{\Omega}\,(y' + \hat{\mu}_2)$.

So $\hat{\mu}_2 \in L(y')$.

(v) Further, $P_{V_1}\,P_{\Omega}\,(y' + \hat{\mu}_2) = P_{V_1}\,\hat{\mu} = P_{\Omega_1} \quad '$.

So $\hat{\mu}_2 \in L^*(y')$.

Thus, we see that $L(y') = L^*(y') =$ the coset $\hat{\mu}_2 + \Omega \cap V_2$.

(vi) Now, $\Omega \cap V_2 = \operatorname{Ker} P_{V_1}\,P_{\Omega}$

$\dim \operatorname{Ker} P_{V_1} \mid \Omega = \dim \Omega - \dim \operatorname{Im} P_{V_1} \mid \Omega$

$= \dim \Omega - \dim \Omega_1$.

Thus, any solution in $L(y')$ will serve equally well to find $P_{\Omega_1} y'$. This generalizes the result of Kruskal [1], where it is assumed that $\dim \Omega = \dim \Omega_1$.

Example:

$$E\, y_{ij} = \alpha_i + \beta_j \qquad i = 1, \cdots, I;\ j = 1, \cdots, J$$

Suppose the i_0, j_0th observation is missing. Then we seek a no. z s.t.

$$\frac{z + y_{\cdot j_0}}{I} + \frac{z + y_{i_0 \cdot}}{J} - \frac{z + y_{\cdot\cdot}}{IJ} = z.$$

Thus,

$$z = \left(\frac{y_{i_0 \cdot}}{J} + \frac{y_{\cdot j_0}}{I} - \frac{y_{\cdot\cdot}}{IJ}\right)\left(\frac{I}{I-1}\right)\left(\frac{J}{J-1}\right).$$

References

[1] William Kruskal, "The Coordinate-Free Approach to Gauss-Markov Estimation, and its Applications to Missing and Extra Observations", Fourth Berkeley Symposium on Math. Stat. and Prob., Vol. 1, pp. 435 - 51.