

Formulas Useful for Linear Regression Analysis and Related Matrix Theory

1 The model matrix & other preliminaries

- 1.1 Linear model. By $\mathcal{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{V}\}$ we mean that we have the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} \in \mathbb{R}^n$ and $\text{cov}(\mathbf{y}) = \sigma^2\mathbf{V}$, i.e., $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $\text{cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{V}$; \mathcal{M} is often called the Gauss–Markov model.
- \mathbf{y} is an observable random vector, $\boldsymbol{\varepsilon}$ is unobservable random error vector, $\mathbf{X} = (\mathbf{1} : \mathbf{X}_0)$ is a given $n \times p$ ($p = k + 1$) model (design) matrix, the vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)' = (\beta_0, \boldsymbol{\beta}'_{\mathbf{x}})'$ and scalar $\sigma^2 > 0$ are unknown
 - $y_i = E(y_i) + \varepsilon_i = \beta_0 + \mathbf{x}'_{(i)}\boldsymbol{\beta}_{\mathbf{x}} + \varepsilon_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \varepsilon_i$, where $\mathbf{x}'_{(i)}$ is the i th row of \mathbf{X}_0
 - from the context it is apparent when \mathbf{X} has full column rank; when distributional properties are considered, we assume that $\mathbf{y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{V})$
 - according to the model, we believe that $E(\mathbf{y}) \in \mathcal{C}(\mathbf{X})$, i.e., $E(\mathbf{y})$ is a linear combination of the columns of \mathbf{X} but we do not know which linear combination
 - from the context it is clear which formulas require that the model has the intercept term β_0 ; p refers to the number of columns of \mathbf{X} and hence in the no-intercept model $p = k$
 - if the explanatory variables x_i are random variables, then the model \mathcal{M} may be interpreted as the conditional model of \mathbf{y} given \mathbf{X} : $E(\mathbf{y} | \mathbf{X}) = \mathbf{X}\boldsymbol{\beta}$, $\text{cov}(\mathbf{y} | \mathbf{X}) = \sigma^2\mathbf{V}$, and the error term is difference $\mathbf{y} - E(\mathbf{y} | \mathbf{X})$. In short, regression is the study of how the conditional distribution of y , when x is given, changes with the value of x .

$$1.2 \quad \mathbf{X} = (\mathbf{1} : \mathbf{X}_0) = (\mathbf{1} : \mathbf{x}_1 : \dots : \mathbf{x}_k) = \begin{pmatrix} \mathbf{x}'_{(1*)} \\ \vdots \\ \mathbf{x}'_{(n*)} \end{pmatrix} \in \mathbb{R}^{n \times (k+1)} \quad \begin{array}{l} \text{model matrix} \\ n \times p, p = k + 1 \end{array}$$

$$1.3 \quad \mathbf{X}_0 = (\mathbf{x}_1 : \dots : \mathbf{x}_k) = \begin{pmatrix} \mathbf{x}'_{(1)} \\ \vdots \\ \mathbf{x}'_{(n)} \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix} \in \mathbb{R}^{n \times k} \quad \begin{array}{l} \text{data matrix} \\ \text{of } x_1, \dots, x_k \end{array}$$

$$1.4 \quad \mathbf{1} = (1, \dots, 1)' \in \mathbb{R}^n, \quad \mathbf{i}_i = (0, \dots, 1 \text{ (} i \text{th)}, \dots, 0)' \in \mathbb{R}^n, \\ \mathbf{i}'_i \mathbf{X} = \mathbf{x}'_{(i*)} = (1, \mathbf{x}'_{(i)}) = \text{the } i \text{th row of } \mathbf{X}, \\ \mathbf{i}'_i \mathbf{X}_0 = \mathbf{x}'_{(i)} = \text{the } i \text{th row of } \mathbf{X}_0$$

$$1.5 \quad \mathbf{x}_1, \dots, \mathbf{x}_k \quad \text{“variable vectors” in “variable space” } \mathbb{R}^n \\ \mathbf{x}_{(1)}, \dots, \mathbf{x}_{(n)} \quad \text{“observation vectors” in “observation space” } \mathbb{R}^k$$

$$1.6 \quad \mathbf{X}_y = (\mathbf{X}_0 : \mathbf{y}) \in \mathbb{R}^{n \times (k+1)} \quad \text{joint data matrix of } x_1, \dots, x_k \text{ and response } y$$

$$1.7 \quad \mathbf{J} = \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}' = \frac{1}{n}\mathbf{1}\mathbf{1}' = \mathbf{P}_1 = \mathbf{J}_n = \text{orthogonal projector onto } \mathcal{C}(\mathbf{1}_n) \\ \mathbf{J}\mathbf{y} = \bar{y}\mathbf{1} = \bar{\mathbf{y}} = (\bar{y}, \bar{y}, \dots, \bar{y})' \in \mathbb{R}^n$$

$$1.8 \quad \mathbf{I} - \mathbf{J} = \mathbf{C} \quad \mathbf{C} = \mathbf{C}_n = \text{orthogonal projector onto } \mathcal{C}(\mathbf{1}_n)^\perp, \\ \text{centering matrix}$$

$$1.9 \quad (\mathbf{I} - \mathbf{J})\mathbf{y} = \mathbf{C}\mathbf{y} = \mathbf{y} - \bar{y}\mathbf{1}_n \\ = \mathbf{y} - \bar{\mathbf{y}} = \tilde{\mathbf{y}} = (y_1 - \bar{y}, \dots, y_n - \bar{y})' \quad \text{centered } \mathbf{y}$$

$$1.10 \quad \bar{\mathbf{x}} = (\bar{x}_1, \dots, \bar{x}_k)' = \frac{1}{n}\mathbf{X}'_0\mathbf{1}_n \\ = \frac{1}{n}(\mathbf{x}_{(1)} + \dots + \mathbf{x}_{(n)}) \in \mathbb{R}^k \quad \text{vector of } x\text{-means}$$

$$1.11 \quad \mathbf{J}\mathbf{X}_y = (\mathbf{J}\mathbf{X}_0 : \mathbf{J}\mathbf{y}) = (\bar{x}_1\mathbf{1} : \dots : \bar{x}_k\mathbf{1} : \bar{y}\mathbf{1}) \\ = (\bar{\mathbf{x}}_1 : \dots : \bar{\mathbf{x}}_k : \bar{\mathbf{y}}) = \mathbf{1}(\bar{\mathbf{x}}', \bar{y}) = \begin{pmatrix} \bar{\mathbf{x}}' & \bar{y} \\ \vdots & \vdots \\ \bar{\mathbf{x}}' & \bar{y} \end{pmatrix} \in \mathbb{R}^{n \times (k+1)}$$

$$1.12 \quad \tilde{\mathbf{X}}_0 = (\mathbf{I} - \mathbf{J})\mathbf{X}_0 = \mathbf{C}\mathbf{X}_0 = (\mathbf{x}_1 - \bar{\mathbf{x}}_1 : \dots : \mathbf{x}_k - \bar{\mathbf{x}}_k) = (\tilde{\mathbf{x}}_1 : \dots : \tilde{\mathbf{x}}_k) \\ = \begin{pmatrix} \mathbf{x}'_{(1)} - \bar{\mathbf{x}}' \\ \vdots \\ \mathbf{x}'_{(n)} - \bar{\mathbf{x}}' \end{pmatrix} = \begin{pmatrix} \tilde{\mathbf{x}}'_{(1)} \\ \vdots \\ \tilde{\mathbf{x}}'_{(n)} \end{pmatrix} \in \mathbb{R}^{n \times k} \quad \text{centered } \mathbf{X}_0$$



<http://www.springer.com/978-3-642-32930-2>

Formulas Useful for Linear Regression Analysis and
Related Matrix Theory

It's Only Formulas But We Like Them

Puntanen, S.; Styan, G.P.H.; Isotalo, J.

2013, XII, 125 p. 3 illus., 2 illus. in color., Softcover

ISBN: 978-3-642-32930-2