Formulas Useful for Linear Regression Analysis and Related Matrix Theory

1 The model matrix & other preliminaries

1.1 Linear model. By $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{V}\}$ we mean that we have the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} \in \mathbb{R}^n$ and $\mathrm{cov}(\mathbf{y}) = \sigma^2\mathbf{V}$, i.e., $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $\mathrm{cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{V}$; $\mathscr{M}$ 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 $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)' = (\beta_0, \boldsymbol{\beta}_{\mathrm{x}}')'$ and scalar $\sigma^2 > 0$ are unknown
- $y_i = \mathrm{E}(y_i) + \varepsilon_i = \beta_0 + \mathbf{x}_{(i)}'\boldsymbol{\beta}_{\mathrm{x}} + \varepsilon_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \varepsilon_i$, where $\mathbf{x}_{(i)}' =$ the ith 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 $\mathbf{y} \sim \mathrm{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{V})$
- according to the model, we believe that $\mathrm{E}(\mathbf{y}) \in \mathscr{C}(\mathbf{X})$, i.e., $\mathrm{E}(\mathbf{y})$ 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 $\mathscr{M}$ may be interpreted as the conditional model of $\mathbf{y}$ given $\mathbf{X}$: $\mathrm{E}(\mathbf{y} \mid \mathbf{X}) = \mathbf{X}\boldsymbol{\beta}$, $\mathrm{cov}(\mathbf{y} \mid \mathbf{X}) = \sigma^2\mathbf{V}$, and the error term is difference $\mathbf{y} - \mathrm{E}(\mathbf{y} \mid \mathbf{X})$. In short, regression is the study of how the conditional distribution of y, when x is given, changes with the value of x.

S. Puntanen et al., *Formulas Useful for Linear Regression Analysis and Related Matrix Theory*, SpringerBriefs in Statistics,
DOI: 10.1007/978-3-642-32931-9_1,

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

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

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

1.5 $\mathbf{x}_1, \ldots, \mathbf{x}_k$ "variable vectors" in "variable space" $\mathbb{R}^n$
$\mathbf{x}_{(1)}, \ldots, \mathbf{x}_{(n)}$ "observation vectors" in "observation space" $\mathbb{R}^k$

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

1.7 $\mathbf{J} = \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}' = \frac{1}{n}\mathbf{1}\mathbf{1}' = \mathbf{P_1} = \mathbf{J}_n =$ orthogonal projector onto $\mathscr{C}(\mathbf{1}_n)$
$\mathbf{J}\mathbf{y} = \bar{y}\mathbf{1} = \bar{\bar{\mathbf{y}}} = (\bar{y}, \bar{y}, \ldots, \bar{y})' \in \mathbb{R}^n$

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

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

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

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

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