Linear Models – Estimation

Consider y_i uncorrelated, $E(y_i) = \mu$, $Var(y_i) = \sigma^2$, $i = 1, 2, \ldots, n$. Estimate μ. In the absence of distributional assumptions, an appealing approach is least squares. What is the estimate and what are its properties? Write the model as:
$y_i = \mu + \epsilon_i$, $E(\epsilon_i) = 0$, $Var(\epsilon_i) = \sigma^2$, $Cov(\epsilon_i, \epsilon_j) = 0$, $i \neq j$. Find

$$\min_{\mu} \sum_{i=1}^{n} (y_i - \mu)^2.$$

Note that,

$$\sum_{i=1}^{n} (y_i - \mu)^2 \quad = \quad \sum_{i=1}^{n} (y_i - \bar{y})^2 + n(\bar{y} - \mu)^2 \geq \sum_{i=1}^{n} (y_i - \bar{y})^2$$

with equality iff $\hat{\mu} = \bar{y}$. Therefore, LSE of μ is $\hat{\mu}_{LS} = \bar{y}$. In vector-matrix formulation,

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \mu \\ \vdots \\ \mu \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix} = \mu\mathbf{1} + \epsilon.$$

$$||\mathbf{Y} - \mu\mathbf{1}||^2 = (\mathbf{Y} - \mu\mathbf{1})\,(\mathbf{Y} - \mu\mathbf{1})' = \sum_{i=1}^{n} (y_i - \mu)^2 = ||\epsilon||^2.$$

Therefore, least squares is equivalent to finding the multiple of $\mathbf{1}$ which minimizes $||\epsilon||$. This is achieved when we take the perpendicular or the orthogonal projection of $\mathbf{Y}$ onto the space spanned by $\mathbf{1}$. i.e.,

$$\frac{\mathbf{Y}'\mathbf{1}}{\mathbf{1}'\mathbf{1}}\mathbf{1} + (\mathbf{Y} - \frac{\mathbf{Y}'\mathbf{1}}{\mathbf{1}'\mathbf{1}}\mathbf{1}) = \mathbf{Y}$$

i.e.,

$$\hat{\mu}_{LS} = \frac{\mathbf{1}'\mathbf{Y}}{\mathbf{1}'\mathbf{1}} = \bar{y}.$$

Since $Cov(\mathbf{Y}) = \sigma^2 I_n$ and $E(\mathbf{Y}) = \mu\mathbf{1}$,

$$E(\hat{\mu}_{LS}) = \frac{1}{\mathbf{1}'\mathbf{1}}\mathbf{1}'E(\mathbf{Y}) = \frac{\mathbf{1}'\mu\mathbf{1}}{\mathbf{1}'\mathbf{1}} = \mu.$$

$$Var(\hat{\mu}_{LS}) \quad = \quad Cov(\frac{\mathbf{1}'\mathbf{Y}}{\mathbf{1}'\mathbf{1}}) = \frac{1}{\mathbf{1}'\mathbf{1}}\mathbf{1}'Cov(\mathbf{Y})\frac{1}{\mathbf{1}'\mathbf{1}}\mathbf{1} = \sigma^2\frac{\mathbf{1}'I_n\mathbf{1}}{(\mathbf{1}'\mathbf{1})^2} = \frac{\sigma^2}{n}.$$

Note, that $\hat{\mu}_{LS}$ is a linear unbiased estimate of μ. Suppose $a'Y$ is any linear unbiased estimate of μ. Then $E(a'Y) = \mu a'1 = \mu$ for all μ implies that $a'1 = 1$. What is the best linear unbiased estimator of μ (i.e., least MSE)? Note,

$$Var(a'Y) = Cov(a'Y) = a'Cov(Y)a = \sigma^2 a'a.$$

To minimize this we just need to find a such that $a'1 = 1$ and $a'a$ is minimum. Simply note that $a'a = \sum_{i=1}^n a_i^2$ and

$$\frac{1}{n}\sum_{i=1}^n a_i^2 - \left(\frac{\sum_{i=1}^n a_i}{n}\right)^2 \geq 0, \text{ for all } a \text{ since } \sum_{i=1}^n (a_i - \bar{a})^2 \geq 0.$$

i.e.,

$$\frac{1}{n}\sum_{i=1}^n a_i^2 - \left(\frac{1}{n}\right)^2 \geq 0, \text{ or } \sum_{i=1}^n a_i^2 \geq \frac{1}{n}$$

with equality iff $a_i = \frac{1}{n}$ for all i. Therefore, $\hat{\mu}_{LS}$ is BLUE (Best Linear Unbiased Estimate) irrespective of the distribution of ϵ.

Linear models: Estimation

Data: $(\mathbf{x}_i, y_i)$, $i = 1, 2, \ldots, n$ with multiple predictors or covariates of y.

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_{p-1} x_{i(p-1)} + \epsilon_i, i = 1, \ldots, n \\ &= \mathbf{x}_i'\beta + \epsilon, i = 1, \ldots, n \end{aligned}$$

is a model for $y|\mathbf{x}$. Let $\mathbf{Y}_{n\times 1} = (y_1, \ldots, y_n)'$, $\beta_{p\times 1} = (\beta_0, \beta_1, \ldots, \beta_{p-1})'$,

$\mathbf{X}_{n\times p} = \begin{pmatrix} x_{10} & x_{11} & \ldots & x_{1(p-1)} \\ \vdots & \vdots & \ldots & \vdots \\ x_{n0} & x_{n1} & \ldots & x_{n(p-1)} \end{pmatrix}$, $x_{i0} \equiv 1$ here but can be general also.

β is called the vector of regression coefficients and $\mathbf{X}$ is called the regression matrix or the design matrix (especially if $x_{ij} = 0$ or 1). Quite often y is called the dependent variable and $\mathbf{x}$ the set of independent variables. It is more standard to call y the response and $\mathbf{x}$, the regressor or predictor. Recall from previous discussion that

$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i$ is a linear model, but

$y_i = \beta_0 + \beta_1 x_i + x_i^{\beta_2} + \epsilon_i$ is nonlinear. i.e., linear model means linear in β_j's.

A general $\mathbf{X}_{n\times p}$ is fine, $\mathbf{X}_0 = \mathbf{1}$ is not essential. Thus we have the linear model:

$\mathbf{Y}_{n\times 1} = \mathbf{X}_{n\times p}\beta_{p\times 1} + \epsilon.$

Since we have only n observations, it does not make sense to consider $p \geq n$, so we take $p < n$. Skip bold face for vectors and matrices unless there is ambiguity.

First task is to estimate β. Most common approach is to use least squares (again, in the absence of distributional assumptions on ϵ). We want

$$\begin{aligned}\min_{\beta\in\mathcal{R}^p}\sum_{i=1}^{n}(y_i - x_i'\beta)^2 &= \min ||\epsilon||^2 = \min_{\beta\in\mathcal{R}^p} ||Y - X\beta||^2 \\ &= \min_{\theta\in\mathcal{M}_C(X)} ||Y-\theta||^2,\end{aligned}$$

where $\mathcal{M}_C(X) = \{a : a = Xb \text{ for some } b \in \mathcal{R}^p\}$. Note that $Xb = b_1X_1 + b_2X_2 + \ldots + b_pX_p$ where X_i are the column vectors of X. Now, to minimize $||Y-\theta||^2$ when $\theta \in \mathcal{M}_C(X)$, we should take $\hat{\theta}$ to be the orthogonal projection of Y onto $\mathcal{M}_C(X)$. i.e., $Y - \hat{\theta}$ should be orthogonal to $\mathcal{M}_C(X)$. i.e.,

$$X'(Y-\hat{\theta}) = 0, \text{ or } X'\hat{\theta} = X'Y.$$

$\hat{\theta}$ is uniquely determined, being the unique orthogonal projection of Y onto $\mathcal{M}_C(X)$. We consider the two cases, Rank$(X) = p$ and Rank$(X) < p$, separately.