

SDS 383D Multiple Regression

Jan-Michael Cabrera, JC7858

February 11, 2019

A: Beta Estimator

Least squares

Least squares involves finding the estimator $\hat{\beta}$ that minimizes the sum of squared errors.

$$\hat{\beta} = \arg \min_{\beta \in \mathcal{R}^p} \left\{ \sum_{i=1}^n (y_i - x_i^T \beta)^2 \right\} \quad (1)$$

The term within the parenthesis can be rewritten in vector notation as

$$f(\beta) = \sum_{i=1}^n (y_i - x_i^T \beta)^2 = (y - x\beta)^T (y - x\beta). \quad (2)$$

This function must then be differentiated and the resultant function set to zero to find the estimator $\hat{\beta}$.

$$\frac{\partial}{\partial \beta} f(\beta) = 0 \quad (3)$$

$$0 = \frac{\partial}{\partial \beta} ((y - x\beta)^T (y - x\beta)) \quad (4)$$

$$= -x^T (y - x\beta) \quad (5)$$

$$= -x^T y + x^T x \beta \quad (6)$$

$$\hat{\beta} = (x^T x)^{-1} x^T y \quad (7)$$

Maximum Likelihood

Maximum likelihood involves finding the estimator $\hat{\beta}$ that maximizes the likelihood function for a given model. Our model is of the form $y_i = x_i^T \beta + \epsilon_i$.

$$\hat{\beta} = \arg \max_{\beta \in \mathcal{R}^p} \left\{ \prod_{i=1}^n p(y_i \mid \beta, \sigma^2) \right\} \quad (8)$$

For our model we assume that the errors are mean zero normally distributed samples each with all the same variances, $\epsilon_i \sim N(0, \sigma^2)$. This gives our error the form $\epsilon_i = y_i - x_i^T \beta$.

The likelihood function is then of the form

$$f(\beta) = \prod_i^n p(y_i | \beta, \sigma^2) \quad (9)$$

$$= \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left[-\frac{n}{2\sigma^2} \sum_i^n (y_i - x_i^T \beta)^2 \right] \quad (10)$$

We can take the log of the above and maximize the resulting function because the logarithm is a monotonically increasing function.

$$\frac{\partial}{\partial \beta} \ln(f(\beta)) = 0 \quad (11)$$

Omitting terms that do not depend on β for brevity,

$$0 = \frac{\partial}{\partial \beta} \left[-\frac{n}{2\sigma^2} \sum_i^n (y_i - x_i^T \beta)^2 + \dots \right] \quad (12)$$

$$= \frac{\partial}{\partial \beta} \left[-\frac{n}{2\sigma^2} (y - x\beta)^T (y - x\beta) \right] \quad (13)$$

$$= -x^T (y - x\beta) \quad (14)$$

$$= -x^T y + x^T x \beta \quad (15)$$

$$\hat{\beta} = (x^T x)^{-1} x^T y \quad (16)$$

Method of Moments

Here we want to choose $\hat{\beta}$ so that the sample covariance between the errors and each predictor is exactly zero, $\text{cov}(\epsilon, x_j) = 0$.

$$0 = \sum_i^n (\epsilon_i - \bar{\epsilon})(x_{ij} - \bar{x}_j) \quad (17)$$

$$= \sum_i^n \epsilon_i x_{ij} - \bar{x}_j \sum_i^n \epsilon_i - \bar{\epsilon} \sum_i^n x_{ij} + \bar{\epsilon} \bar{x}_j \quad (18)$$

We observe that $\bar{\epsilon} = \frac{1}{n} \sum_i^n \epsilon_i$ and $\bar{x}_j = \frac{1}{n} \sum_i^n x_{ij}$. We can also center our data such that $\bar{x}_j = 0$.

$$0 = \sum_i^n \epsilon_i x_{ij} \tag{19}$$

$$= \epsilon^T x \tag{20}$$

$$= (y - x\beta)^T x \tag{21}$$

$$= y^T x - \beta^T x^T x \tag{22}$$

$$= x^T y - x^T x \beta \quad \text{with} \quad \beta^T x^T x = x^T x \beta \quad \text{and} \quad y^T x = x^T y \tag{23}$$

$$\tag{24}$$

$$\hat{\beta} = (x^T x)^{-1} x^T y \tag{25}$$