

# CS&SS/STAT/SOC 536: Linear Regression

Adrian Dobra  
adobra@uw.edu

## 1 Introduction

We assume the simplest case of one response (independent) variable  $y$  and one explanatory (dependent) variable  $x$ :

$$y = \beta_0 + \beta_1 x + \epsilon. \quad (1)$$

We assume that the errors are normally distributed and homoscedastic (i.e., they are constant for all observed samples; antonym: heteroscedastic):

$$\epsilon \sim N(0, \sigma^2), \quad \sigma > 0.$$

This implies that:

$$\begin{aligned} y &\sim N(\beta_0 + \beta_1 x, \sigma^2), \\ E[y|x] &= \beta_0 + \beta_1 x, \\ \text{Var}[y|x] &= \sigma^2. \end{aligned}$$

The linear regression model (1) has three parameters  $\beta_0$ ,  $\beta_1$  and  $\sigma$  that will be estimated from the available data. Denote these estimates by  $\hat{\beta}_0$ ,  $\hat{\beta}_1$  and  $\hat{\sigma}$ . We usually denote the model parameters by  $\theta$  and their estimates by  $\hat{\theta}$ . In this particular case, we have

$$\theta = (\beta_0, \beta_1, \sigma), \quad \hat{\theta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}).$$

The predicted value  $y_0$  of the response given some value  $x_0$  of the explanatory variable is:

$$y_0 \sim N(\hat{\beta}_0 + \hat{\beta}_1 x_0, \hat{\sigma}^2).$$

The interpretation of the regression coefficients depends on the nature of the explanatory variable.

- $x$  is continuous (any real value).

$$\frac{\partial E[y|x]}{\partial x} = \hat{\beta}_1.$$

- $x$  is binary (i.e.,  $x \in \{0, 1\}$ ).

$$E[y|x = 1] - E[y|x = 0] = \hat{\beta}_1.$$

- $x$  is a transformed version of another variable  $z$ , i.e.  $x = h(z)$  with  $h \equiv \log(\cdot)$ ,  $h \equiv \sqrt{\cdot}$  or  $h \equiv \exp(\cdot)$ .

## 2 Maximum Likelihood Estimation

Let  $(y_1, x_1), \dots, (y_n, x_n)$  denote the observed data. Given the linear regression model (1), we have:

$$y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2), \quad \text{for } i = 1, \dots, n. \quad (2)$$

In matrix notation, we write:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 & \beta_1 \end{bmatrix}^T + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix},$$

or, in a more concise form:

$$Y = X\beta + \epsilon.$$

The density of the standard normal  $N(0, 1)$  is  $\phi(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2)$ . Its derivative is  $\phi'(z) = -z\phi(z)$ . The likelihood associated with (2) is written as:

$$L(\theta) = L(\beta_0, \beta_1, \sigma) = \prod_{i=1}^n \frac{1}{\sigma} \phi\left(\frac{y_i - (\beta_0 + \beta_1 x_i)}{\sigma}\right).$$

The log-likelihood (i.e., the natural logarithm of the likelihood) is

$$l(\theta) = l(\beta_0, \beta_1, \sigma) = -n \log \sigma + \sum_{i=1}^n \log \phi\left(\frac{y_i - (\beta_0 + \beta_1 x_i)}{\sigma}\right).$$

The MLE  $\hat{\theta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma})$  is obtained by solving the system of equations:

$$l'(\theta) = 0,$$

or, equivalently

$$\frac{\partial l(\theta)}{\partial \sigma} = 0, \quad \frac{\partial l(\theta)}{\partial \beta_0} = 0, \quad \frac{\partial l(\theta)}{\partial \beta_1} = 0.$$

We have

$$\frac{\partial l(\theta)}{\partial \sigma} = -\frac{n}{\sigma} + \sum_{i=1}^n \left(\frac{y_i - (\beta_0 + \beta_1 x_i)}{\sigma}\right)^2 \frac{1}{\sigma}.$$

By solving  $\frac{\partial l(\theta)}{\partial \sigma} = 0$ , we obtain the MLE of  $\sigma$ :

$$\hat{\sigma} = \left[ \frac{1}{n} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2 \right]^{1/2}. \quad (3)$$

Notice that  $\hat{\sigma}$  is a function of the estimates of  $\beta_0$  and  $\beta_1$ , hence we will need them before being able to evaluate  $\hat{\sigma}$ . The equation  $\frac{\partial l(\theta)}{\partial \beta_0} = 0$  is:

$$\sum_{i=1}^n y_i = n\beta_0 + \left( \sum_{i=1}^n x_i \right) \beta_1. \quad (4)$$

Similarly, the equation  $\frac{\partial l(\theta)}{\partial \beta_1} = 0$  is:

$$\sum_{i=1}^n x_i y_i = \left( \sum_{i=1}^n x_i \right) \beta_0 + \left( \sum_{i=1}^n x_i^2 \right) \beta_1. \quad (5)$$

Equations (4) and (5) form a system of two linear equations with two unknowns, which has a unique solution  $(\hat{\beta}_0, \hat{\beta}_1)$ . These are the MLEs of  $\beta_0$  and  $\beta_1$ . We substitute  $(\hat{\beta}_0, \hat{\beta}_1)$  in (3) to obtain the MLE of  $\sigma$ .

One way to check our derivation is to consider the familiar OLS estimator:

$$\hat{\beta} = (X^T X)^{-1} X^T y \Leftrightarrow (X^T X) \hat{\beta} = X^T y.$$

Simple matrix multiplication shows that we obtain precisely the equations (4) and (5).

Under certain regularity conditions, the asymptotic covariance of the MLEs is the inverse of the Hessian of  $l(\theta)$ , evaluated at the MLEs (SEE YOUR HANDOUT “LIKELIHOOD INFERENCE”):

$$Cov(\hat{\theta}) \approx -H^{-1}(\hat{\theta}).$$

We have

$$H(\theta) = H(\beta_0, \beta_1, \sigma) = \begin{bmatrix} \frac{\partial^2 l(\theta)}{\partial \beta_0^2} & \frac{\partial^2 l(\theta)}{\partial \beta_0 \partial \beta_1} & \frac{\partial^2 l(\theta)}{\partial \beta_0 \partial \sigma} \\ \frac{\partial^2 l(\theta)}{\partial \beta_1 \partial \beta_0} & \frac{\partial^2 l(\theta)}{\partial \beta_1^2} & \frac{\partial^2 l(\theta)}{\partial \beta_1 \partial \sigma} \\ \frac{\partial^2 l(\theta)}{\partial \sigma \partial \beta_0} & \frac{\partial^2 l(\theta)}{\partial \sigma \partial \beta_1} & \frac{\partial^2 l(\theta)}{\partial \sigma^2} \end{bmatrix}.$$

Simple calculations show that  $\frac{\partial^2 l(\theta)}{\partial \beta_0^2} = -\frac{n}{\sigma^2}$ ,  $\frac{\partial^2 l(\theta)}{\partial \beta_0 \partial \beta_1} = -\frac{1}{\sigma^2} \left( \sum_{i=1}^n x_i \right)$ , etc...