Simple Linear Regression Model – Part V

Johns Hopkins Engineering

625.461 Statistical Models and Regression

Module 2 – Lecture 2D

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Estimation by Maximum Likelihood

If $y_i \mid x_i \sim \mathrm{NIND}(\beta_0 + \beta_1 x_i\,,\ \sigma^2),$

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \ , \qquad i = 1, \ldots, n$$

then we can write down likelihood function of n observations

Estimation by Maximum Likelihood

$$L(y_i, x_i, \beta_0, \beta_i, \sigma^2) = \prod_{i=1}^{n} (2\pi\sigma^2)^{-1/2} \exp\left[-\frac{1}{2\sigma^2}(y_i - \beta_0 - \beta_1 x_i)^2\right]$$

$$= (2\pi\sigma^2)^{-1/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2\right]$$

Maximum likelihood estimators (MLE) are the parameter values to maximize *L* or equivalently *ln L*

Estimation by Maximum Likelihood

$$\ln L\left(y_i, x_i, \beta_0, \beta_i, \sigma^2\right) = -\left(\frac{n}{2}\right)\ln 2\pi - \left(\frac{n}{2}\right)\ln \sigma^2$$

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

MLEs $\tilde{\beta}_0$, $\tilde{\beta}_1$, and $\tilde{\sigma}^2$ must satisfy

This image cannot currently be displayed.

Estimation by Maximum Likelihood

$$\left.\frac{\partial \ln L}{\partial \sigma^2}\right|_{\tilde\beta_0,\tilde\beta_1,\tilde\sigma^2} = -\frac{n}{2\tilde\sigma^2} + \frac{1}{2\tilde\sigma^4}\sum_{i=1}^{n}\left(y_i - \tilde\beta_0 - \tilde\beta_1 x_i\right)^2 = 0$$

Consequently, MLEs are:

$$\tilde\beta_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad \tilde\beta_0 = \bar{y} - \tilde\beta_1\bar{x} \qquad \tilde\sigma^2 = \frac{\sum_{i=1}^{n}\left(y_i - \tilde\beta_0 - \tilde\beta_1 x_i\right)^2}{n}$$

Estimation by Maximum Likelihood

For β_0 and β_1 , MLE = OLSE

For σ^2 , MLE $\neq$ OLSE

$$\tilde{\sigma}^2 = \left[{(n-2)}\big/{n} \right] \hat{\sigma}^2$$

MLE is biased for σ^2, but the bias is smaller as n is larger

Regressor is a Random Variable

Suppose that y and x are random variables with unknown joint distribution.

All of our previous regression results still hold if:

- $y \mid x \sim N(\beta_0 + \beta_1 x\,,\, \sigma^2)$
- x's independent random variables with probability distribution not involving $(\beta_0\,,\, \beta_1\,,\, \sigma^2)$

Random Regressor

If y and x are jointly bivariate normal (correlation model), the joint density is

$$f(y,x)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

$$\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y-\mu_1}{\sigma_1}\right)^2+\left(\frac{x-\mu_2}{\sigma_2}\right)^2-2\rho\left(\frac{y-\mu_1}{\sigma_1}\right)\left(\frac{x-\mu_2}{\sigma_2}\right)\right]\right\}$$

$$\rho=\frac{E(y-\mu_1)(x-\mu_2)}{\sigma_1\sigma_2}=\frac{\sigma_{12}}{\sigma_1\sigma_2}$$:correlation between y & x

$$E(y|x)=\beta_0+\beta_1 x \qquad Var(y|x)=\sigma_1^2(1-\rho^2)$$

Between Slope Estimator and Sample Correlation Coefficient

MLE for β_0 and β_1 remain the same.

Sample correlation coefficient:

$$r = \frac{\sum_{i=1}^{n} y_i (x_i - \bar{x})}{\left[\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2\right]^{1/2}} = \frac{S_{xy}}{[S_{xx} SS_{\mathrm{T}}]^{1/2}}$$

$$\hat{\beta}_1 = \left(\frac{SS_{\mathrm{T}}}{S_{xx}}\right)^{1/2} r$$

Sample Correlation Coefficient

$$r^2 = \hat{\beta}_1^2 \frac{S_{xx}}{SS_T} = \frac{\hat{\beta}_1 S_{xy}}{SS_T} = \frac{SS_R}{SS_T} = R^2$$

R^2 is coefficient of determination

Test for Correlation Coefficient

To test $H_0: \rho = 0, \quad H_1: \rho \neq 0$,
use the test

$$t_0 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

which is t distribution with df = $n - 2$
under $H_0: \rho = 0$.

JOHNS HOPKINS

WHITING SCHOOL *of* ENGINEERING