The expected value and variance of the estimators in simple linear regression

(STATS310.3: Simple linear regression)

Gunnar Stefansson

June 9, 2012

Expected value of the slope estimator

The estimator for the slope is unbiased:

$$\hat{\beta} = \sum_i \frac{(x_i - \bar{x})Y_i}{\sum_j (x_j - \bar{x})^2}$$

$$\Rightarrow E\hat{\beta} = \sum_i \frac{(x_i - \bar{x})E[Y_i]}{\sum_j (x_j - \bar{x})^2}$$

$$= \sum_i \frac{(x_i - \bar{x})(\alpha + \beta x_i)}{\sum_j (x_j - \bar{x})^2} = \ldots = \beta$$

Note: This result only depends on the mean structure of Y_i, not the p.d.f. or even the variance.

Variance of the slope estimator

The variance of the estimator can be derived:

$$V\left[\hat{\beta}\right] = \ldots = \frac{\sigma^2}{\sum(x - \bar{x})^2}$$

Note: This result only depends on the mean and variance structure of Y_i, not the p.d.f.

Expected value of the intercept estimator

The estimate of the intercept is unbiased:

$$\begin{aligned} E\hat{\alpha} &= E\left[\bar{Y} - \hat{\beta}\bar{x}\right] \\ &= E\left[\bar{Y}\right] - \beta\bar{x} \\ &= (\alpha + \beta\bar{x}) - \beta\bar{x} \\ &= \alpha. \end{aligned}$$

Note: This result only depends on the mean and variance structure of Y_i, not the p.d.f.

Variance of intercept estimator

The variance of the estimator can be derived:

$$V[\hat{\alpha}] = \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right)$$

Note: This result only depends on the mean and variance structure of Y_i, not the p.d.f.

Estimating slope accuracy

The standard error of the slope:

$$\hat{\sigma}^2_{\hat{\beta}} = \frac{\hat{\sigma}^2}{\sum(x - \bar{x})^2}$$

where

$$\hat{\sigma}^2 = \frac{\sum(y - \hat{y})^2}{n - 2}$$

Experimental design issues

The formulae for variances of slope and intercept can be used to obtain optimal design

Would like $\bar{x}$ close to 0

Ideally dispersion of x-values should be large