Ch 3: Linear Regression

Alex Socarras

3/29/2022

SLR: $Y \approx \beta_0 + \beta_1 X, \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$

RSS: $RSS = e_1^2 + e_2^2 + \cdots + e_n^2,$ where $e_i = y_i - \hat{y}_i$

LSCE: $\hat{\beta}_1 = \dfrac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$ and $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

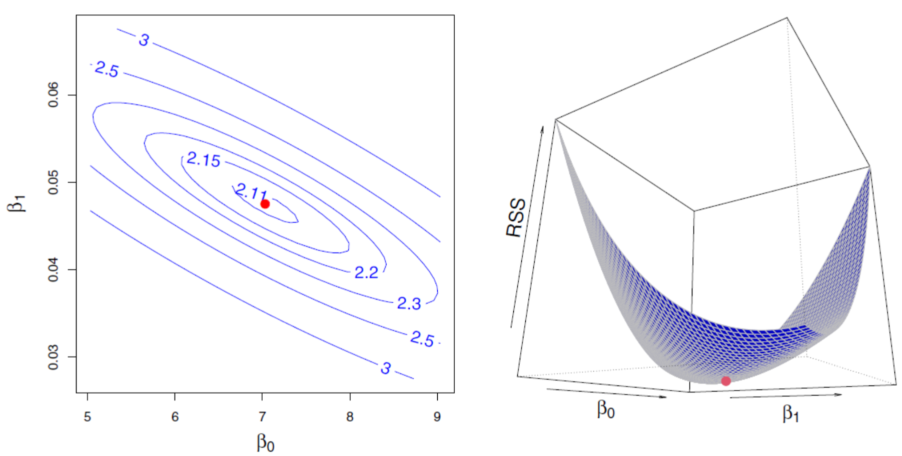


FIGURE 3.2. *Contour and three-dimensional plots of the RSS on the* `Advertising` *data, using* `sales` *as the response and* `TV` *as the predictor. The red dots correspond to the least squares estimates $\hat{\beta}_0$ and $\hat{\beta}_1$, given by (3.4).*

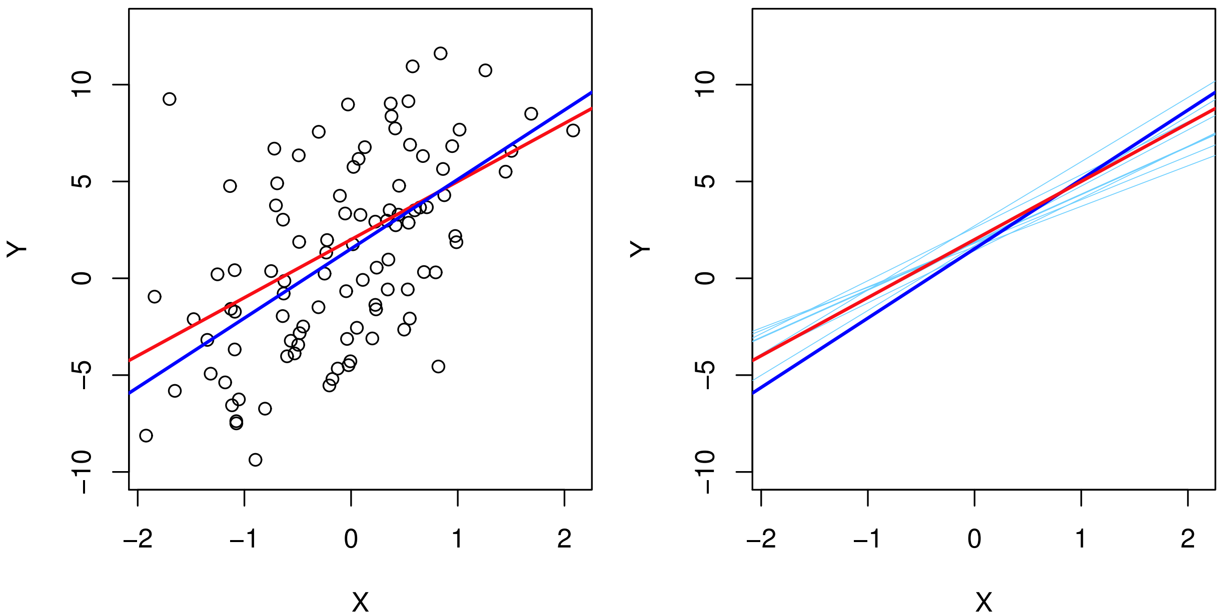
Model Accuracy

population regression line: $Y = \beta_0 + \beta_1 X + \epsilon$ (assume ϵ is ind. of X)

Estimating β_0, β_1 is analogous to estimating the population mean of a random variable by averaging estimates obtained over a large number of data sets (samples):

$$\text{Var}(\hat{\mu}) = \text{SE}(\hat{\mu})^2 = \frac{\sigma^2}{n},$$

where the *standard error* of $\hat{\mu} = SE(\hat{\mu})$ indicates the average amount the estimated mean differs from the population mean. **Warning:** assumes n are uncorrelated.



Y
X
Y
X

•