STAT 410 - Linear Regression

Lecture 2

Meng Li

Department of Statistics

Aug. 24, 2017

Simple linear regression model

- Simple linear regression model:

$$y = \beta_0 + \beta_1 x + \varepsilon, \tag{1}$$

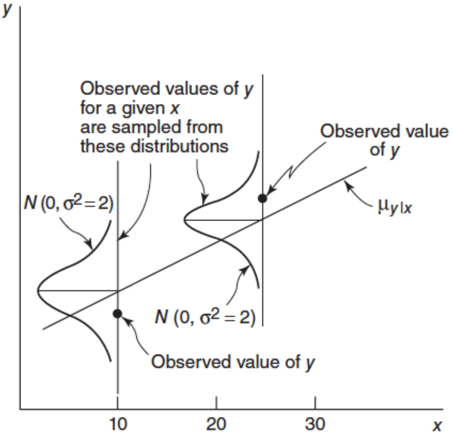
  where ε has mean 0 and $\sigma^2 < \infty$.
- Question: terminology associated with x, y, β_0, β_1.
- SLR implies

$$\mathrm{E}(y|x) = \beta_0 + \beta_1 x, \quad \mathrm{Var}(y|x) = \sigma^2.$$

- Thus, given x, the mean of y is a linear function of x but the variance of y does not depend on x.
- Interpretation of β_1: change in the mean of the distribution of y produced by a unit change in x.
- Interpretation of β_0: the mean of the distribution of the response y when x = 0.

Simple linear regression model with normal errors

- We further *assume* that ε is Gaussian, i.e., $\varepsilon \sim N(0, \sigma^2)$.
- This means given x the response is generated by $N(\beta_0 + \beta_1 x, \sigma^2)$.



Estimation of parameters

- Let our data (or observations) be $(x_1, y_1), \ldots, (x_n, y_n)$
- Then SLR model assumes

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

  where $\varepsilon_i \overset{\text{iid}}{\sim} \text{N}\left(0, \sigma^2\right)$.
- Parameters in the model: $(\beta_0, \beta_1, \sigma^2)$
- In STAT 310, you might deal with the model $y_i = \mu + \varepsilon_i$, or $y_i \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$.
- We here use the maximum likelihood estimators (MLE).
- The likelihood function is

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

- Then the MLE is obtained by

$$(\hat{\beta}_0^{MLE}, \hat{\beta}_1^{MLE}, \widehat{\sigma^2}_{MLE}) = \underset{\beta_0, \beta_1, \sigma^2}{\operatorname{argmax}}\, L(\beta_0, \beta_1, \sigma^2).$$

MLE of (β_0, β_1) - a sketch

- The log-likelihood function simplifies to

$$\log L(\beta_0, \beta_1, \sigma^2) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 + \text{constant}$$

- MLE of (β_0, β_1) satisfies that

$$(\hat\beta_0^{MLE}, \hat\beta_1^{MLE}) = \underset{\beta_0,\beta_1}{\operatorname{argmin}} \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2.$$

1. Set the first derivative to be 0.

$$\begin{cases} \beta_0 \sum_{i=1}^n x_i + \beta_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i \\ \beta_0 n + \beta_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \end{cases}$$

2. Solve equations. Let $\bar{x} = \sum_{i=1}^n x_i/n$ and $\bar{y} = \sum_{i=1}^n y_i/n$.

$$\hat\beta_1^{MLE} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad (2)$$

$$\hat\beta_0^{MLE} = \bar{y} - \hat\beta_1^{MLE}\bar{x} \quad (3)$$

3. Check the 2nd derivative - negative at $(\hat\beta_0^{MLE}, \hat\beta_1^{MLE})$.

MLE of σ^2

- Similarly, we have
$$\widehat{\sigma^2}_{MLE} = \frac{\sum_{i=1}^{n}\{y_i - (\hat{\beta}_0^{MLE} + \hat{\beta}_1^{MLE} x_i)\}^2}{n}. \qquad (4)$$
- Notation:
$$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}(x_i - \bar{x})\,x_i = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$$
$$S_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \bar{y})\,y_i = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2$$
$$S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})\,(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y};$$
Fitted values: $\hat{y}_i = \hat{\beta}_0^{MLE} + \hat{\beta}_1^{MLE} x_i$

Residues: $e_i = y_i - \hat{y}_i$
- We then rewrite the MLE in equations (2), (3) and (4) to
$$\hat{\beta}_1^{MLE} = S_{xy}/S_{xx}, \quad \hat{\beta}_0^{MLE} = \bar{y} - \hat{\beta}_1^{MLE}\bar{x}, \quad \widehat{\sigma^2}_{MLE} = \sum_{i=1}^{n} e_i^2/n.$$

Least-squares estimators of (β_0, β_1)

- Least squares estimation seeks to minimize the sum of squares of the differences between the observed response, y_i, and the straight line.
- $$(\hat{\beta}_0^{LS}, \hat{\beta}_1^{LS}) = \underset{\beta_0, \beta_1}{\operatorname{argmin}} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2.$$
- Therefore, least-square estimators of (β_0, β_1) are identical to MLE's under normal assumption, i.e.,

$$\hat{\beta}_1^{LS} = S_{xy}/S_{xx}, \quad \hat{\beta}_0^{LS} = \bar{y} - \hat{\beta}_1^{LS}\bar{x}.$$

- We simply use $(\hat{\beta}_0, \hat{\beta}_1)$ for both maximum likelihood and (ordinary) least-square estimators.

Unbiased estimator of σ^2

- Sample variances of $e_1, \ldots, e_n$ is a sensible estimator of σ^2
- It turns out that

$$\widehat{\sigma^2} = \frac{\sum_{i=1}^n \{y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)\}^2}{n-2} = \frac{\sum_{i=1}^n e_i^2}{n-2} \tag{5}$$

  is an unbiased estimator of σ^2.
- This is different from $\hat{\sigma}^2_{MLE}$!

MLE vs. OLS

- (Ordinary) least-square estimators do not require normal assumption but relies on a loss function.
- Least-square estimators certainly change if we switch the loss function.
- They are identical for both the intercept and slope.
- MLE provides a statistical justification to least-square estimator on its use of squared loss.

Computation

- Use the previous Delivery example
- The least-square line is

$$\hat{y} = 3.321 + 2.1762x$$

- `R` script “delivery.R" available online to demonstrate the calculation.

Discussion & What's next

- Other estimators do exist, for example, Bayesian estimators.
- It thus becomes necessary to investigate properties of an estimator to allow comparison.
- Furthermore, we would like to draw inferences beyond a point estimate.
- These discussions will be covered by the next lectures.