Simple linear regression: II. point estimation

Miaoyan Wang

Department of Statistics
UW Madison

Simple linear regression

References:

- Chapter 2 in JF (Julian J. Faraway)
- Chapter 2.1-2.9, 2.11 in RC (Ronald Christensen)

Both textbooks are available in Canvas/files/textbook/

Recall: simple linear regression model

- A straight line relationship between the response variable Y and the explanatory variable X:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad \text{where} \quad E(\varepsilon_i) = 0$$

- Equal variance:

$$Var(\varepsilon_i) = \sigma^2.$$

- Independence:

$$Cov(\varepsilon_i, \varepsilon_{i'}) = 0 \quad \text{for} \quad i \neq i'.$$

- Normal distribution:

$$\varepsilon_i \sim N(0, \sigma^2).$$

Q: ε_i are iid. How about Y_i? iid? Not iid? It depends?

Model Parameters

- The model parameters are β_0, β_1, and σ^2 (population parameters).
- β_0 and β_1: **regression coefficients**.
- β_0: **intercept**.
  When the model scope includes $x = 0$, β_0 can be interpreted as the mean of Y at $x = 0$.
- β_1: **slope**.
  Interpreted as the change in the mean of Y per unit increase in x.
- σ^2: **error variance**, sometimes written as σ_ε^2 or $\sigma_{Y|x}^2$.

Q: How to estimate the model parameters based on data?

Estimation of Model Parameters

- Our goal is to estimate these model parameters by estimators $\hat{\beta}_0, \hat{\beta}_1$, and $\hat{\sigma}^2$, based on data.
- Two methods:
  - Least squares (LS).
  - Maximum likelihood (ML).
- Additional notation:
  - Let $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ denote the ith fitted value.
  - Let $e_i = Y_i - \hat{Y}_i$ denote the ith residual.

Estimation of β_0 and β_1

- Both LS and ML give the same estimator for β_0 and β_1:

$$
\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \\
\hat{\beta}_0 &= \frac{1}{n}\left(\sum_{i=1}^{n} Y_i - \hat{\beta}_1 \sum_{i=1}^{n} X_i\right) = \bar{Y} - \hat{\beta}_1 \bar{X}.
\end{aligned}
$$

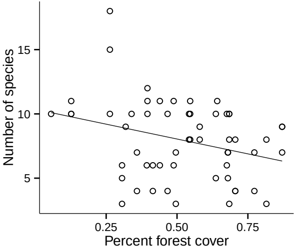
- We now give **two methods** for these estimations.

Least Squares (LS) Estimation

- Consider the criterion:

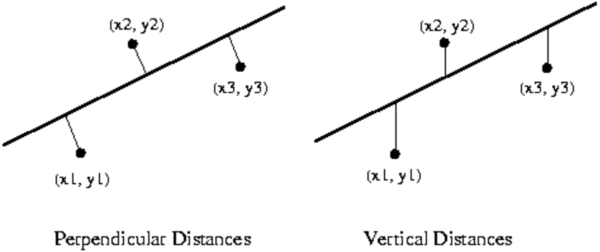
$$Q = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2.$$

- The LS estimators of β_0 and β_1 are those values, $\hat{\beta}_0$ and $\hat{\beta}_1$, that minimize Q, for the given observed data $(X_1, Y_1), \ldots, (X_n, Y_n)$.
- Graphical interpretation?



Brainstorm

Why do we use vertical distance to define the fitted line?



Other choices?

- The sum of the squares of perpendicular distance
- The sum of absolute value of the distance

Approach 1: LS Derivation

- Differentiate Q with respect to β_0 and β_1:

$$
\begin{aligned}
\text{(a)} \quad &: \quad \frac{\partial Q}{\partial \beta_0} = -2 \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i) \\
\text{(b)} \quad &: \quad \frac{\partial Q}{\partial \beta_1} = -2 \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i) X_i
\end{aligned}
$$

- Set (a) and (b) equal to 0 and let the solutions to these two equations be $\hat{\beta}_0$ and $\hat{\beta}_1$.
- Let $\boldsymbol{\beta} = (\beta_0, \beta_1)'$.
- Since $\frac{\partial^2 Q}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}$ is positive definite, $\hat{\beta}_0$ and $\hat{\beta}_1$ minimize Q.

Approach 2: ML Derivation

- Let $\boldsymbol{\theta} = (\beta_0, \beta_1, \sigma^2)'$.
- We have $Y_i \sim \text{i.i.d.} N(\beta_0 + \beta_1 X_i, \sigma^2)$.
- Thus,

$$f_i(y_i; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\left\{y_i - (\beta_0 + \beta_1 x_i)\right\}^2\right].$$

- The likelihood function is

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y}) &= \prod_{i=1}^{n} f_i(y_i; \boldsymbol{\theta}) \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\left\{y_i - (\beta_0 + \beta_1 x_i)\right\}^2\right]
\end{aligned}$$

ML Derivation (Cont.)

Solve for the parameters and obtain the ML estimates:

$$
\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} \\
\hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X} \\
\tilde{\sigma}^2 &= \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}
\end{aligned}
$$

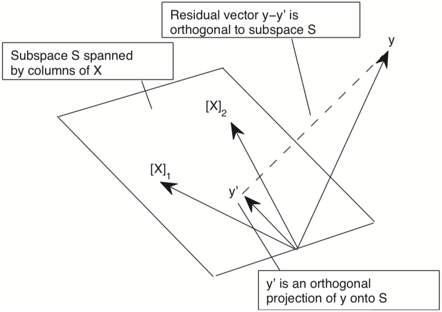
Properties of Fitted Regression Line

For the fitted values $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ and residuals $e_i = Y_i - \hat{Y}_i$, we have:

- The regression line always goes through $(\bar{X}, \bar{Y})$.
- $\sum_{i=1}^{n} e_i^2$ is a minimum.
- $\sum_{i=1}^{n} e_i = 0$.
- $\sum_{i=1}^{n} X_i e_i = 0$.
- $\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \hat{Y}_i$.
- $\sum_{i=1}^{n} \hat{Y}_i e_i = 0$.

Exercises: Proofs of the above properties.

Geometric Interpretation



- Define "hat matrix" (projection matrix): $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ and $\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\beta} = \boldsymbol{H}\boldsymbol{Y}$
- $\boldsymbol{H}$ projects $\boldsymbol{Y}$ onto the span of $\boldsymbol{X}$.
- $\boldsymbol{I} - \boldsymbol{H}$ projects $\boldsymbol{Y}$ onto the space orthogonal to $\boldsymbol{X}$.
- Exercise: What is the algebraic properties of the hat matrix $\boldsymbol{H}$?
  rank, eigenvalues-vectors, semi positive definite, idempotent, etc.

Estimation of σ^2

- Define an **error sum of squares (SSE)** (or, **residual sum of squares**):

$$\text{SSE} = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} e_i^2.$$

- Under simple linear regression, an unbiased estimate of σ^2 is an **error mean square (MSE)** (or, **residual mean square**):

$$\hat{\sigma}^2 = \text{MSE} = \frac{\text{SSE}}{n-2} = \frac{\sum_{i=1}^{n} e_i^2}{n-2}.$$

- The biased ML estimate of σ^2 is:

$$\tilde{\sigma}^2 = \frac{\text{SSE}}{n} = \frac{\sum_{i=1}^{n} e_i^2}{n}.$$

Example: Wetland Species Richness

- In the wetland species richness example, we have

$$\text{SSE} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2 = 479.04$$

- Under LS, we have

$$\hat{\sigma}^2 = \text{MSE} = \frac{\text{SSE}}{n-2} = \frac{479.04}{56} = 8.554$$

- Under ML, we have

$$\tilde{\sigma}^2 = \frac{\text{SSE}}{n} = \frac{479.04}{58} = 8.259.$$

- Which estimator is better?

Q: Why df $= n - 2$ for MSE? Which estimator is better?
Note: $E(\tilde{\sigma}^2) = \frac{n-2}{n}\sigma^2$.