P8130: Biostatistical Methods I

Lecture 13: Simple Linear Regression

Cody Chiuzan, PhD
Department of Biostatistics
Mailman School of Public Health (MSPH)

Outline

- Lecture 12 introduced (simple) liner regression (SLR)
- Least squares estimators (LS)

- Today we will discuss:
  - Properties of LS estimators
  - Maximum likelihood estimators (MLEs) and properties
  - Matrix notation for linear regression

LS Estimation

- Given the linear model: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$
  - Minimize criterion $Q = \sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)^2$ to obtain the LS estimates:

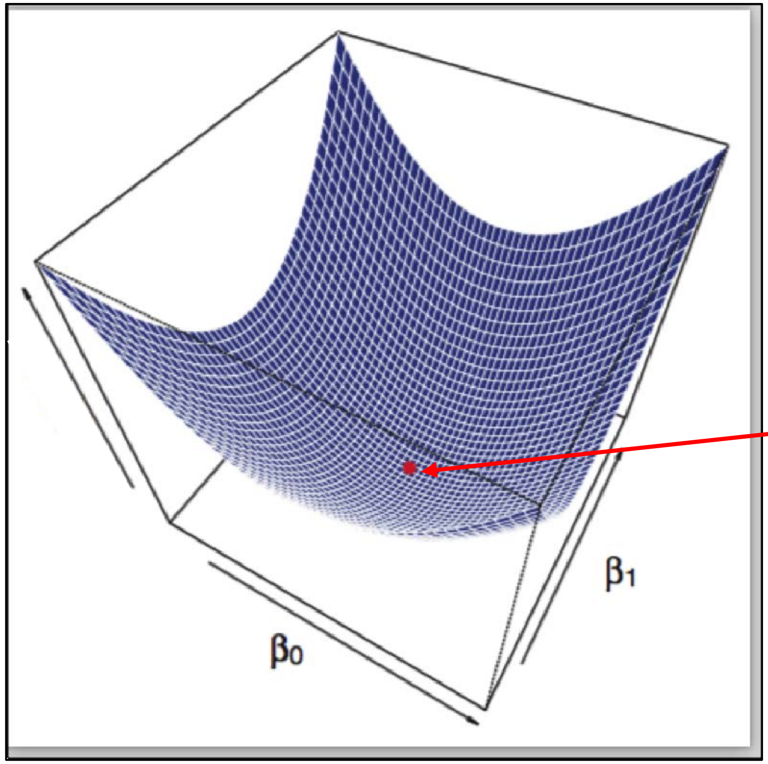
$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}}{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}$$

Residual Variance Estimation

- Sum of square errors: $SSE = \sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2 = \sum_{i=1}^{n} e_i^2$

MSE (mean square error): $s^2 = MSE = \frac{SSE}{n-2} = \frac{\sum_{i=1}^{n} e_i^2}{n-2}$



Example of a 3D-plot of SSE for different parameter values.

The red dot is the pair of LS estimates representing the local and absolute minimum: $(\widehat{\beta_0}, \widehat{\beta_1})$.

Properties of LS Estimators

- Simply mathematical, but closely identified with the Gaussian error model ML estimators

- $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased for β_0 and β_1, respectively (show in class)

$$E(\hat{\beta}_0) = \beta_0$$

$$E(\hat{\beta}_1) = \beta_1$$

- Note that *MSE* is also an unbiased estimator of the error variance:

$$E(MSE) = E\left(\frac{SSE}{n-2}\right) = E\left(\frac{\sum_{i=1}^{n} e_i^2}{n-2}\right) = \sigma^2$$

Add ‘Probability’ to our model

- So far we only assumed that $E(\varepsilon_i) = 0$ and $\sigma^2(\varepsilon_i) = \sigma^2$
- Residuals are *i.i.d.,* independent and identically distributed (following a normal distribution):

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

- This new assumption allows us to make *inferences* about the model parameters and obtain prediction intervals for a new Y_{n+1}

Add 'Probability' to our model

- Succinctly, the model is:

$$Y_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2), i = 1, 2, \ldots, n$$

- Recall the likelihood of a normal distribution:

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

- Given that $E(e_i) = \mu = 0$, the likelihood of the linear model becomes:

$$L(\mu, \sigma^2 \mid x) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(Y_i - \beta_0 - \beta_1 X_i)^2}{2\sigma^2}\right)$$

Maximizing the likelihood of the SLR

- Maximize the log-likelihood (score function):

$$\ln L(\mu,\sigma^2 \mid x) = \log\left[\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(Y_i - \beta_0 - \beta_1 X_i)^2}{2\sigma^2}\right)\right]$$

...

$$\ln L(\mu,\sigma^2 \mid x) = -\frac{n}{2}\log(2\pi) - n\log(\sigma) - \sum_{i=1}^{n} \frac{(Y_i - \beta_0 - \beta_1 X_i)^2}{2\sigma^2}$$

- Take the derivatives with respect to each of the two parameters, equal to 0 and solve the system of equations.

- Surprise!
  - The maximum likelihood estimators (MLE) for parameters are the same as via LS estimation.

ML vs LS Estimation

- For linear regression, the two methods of estimation give similar results.

  However, for other regressions, e.g., Logistic, Poisson, the estimates differ

- Remember, LS requires no error distribution assumption, but ML does

- For inferences (and to avoid confusion), from this point forward we will assume/use normal residuals regression model

Properties of ML Estimators

- As LS, the ML estimators are (asymptotically) unbiased:

$$E(\hat{\beta}_0) = \beta_0, \text{ at least as n} \to \infty$$

$$E(\hat{\beta}_1) = \beta_1$$

- But the variance estimator $\hat{\sigma}^2_{MLE} = \sum_{i=1}^{n} \hat{\varepsilon}_i^2$ is only unbiased as $n \to \infty$.

- What about software results?
  - Beta(s) are the same for LS ad MLE
  - Variance estimator? – lm() function in R provides the unbiased estimator (residual standard error)

Inferences about parameters

- To make inferences we need to understand the sampling distribution

- Remember that $\hat{\beta}_0, \hat{\beta}_1$ are linear combinations of Y's, which are normal random variables

- Thus, $\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right)$

- It follows that $\frac{\widehat{\beta}_1 - \beta_1}{\sqrt{\frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}} \sim N(0,1) \rightarrow \frac{\widehat{\beta}_1 - \beta_1}{\sqrt{\frac{MSE}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}} \sim t_{n-2}$

Inferences about parameters

- Same idea for β_0.
- If $\hat{\beta}_0 \sim N\left(\beta_0, \sigma^2\left(\frac{1}{n}+\frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right)\right)$ then it follows that:

$$\frac{\hat{\beta}_0-\beta_0}{\sqrt{MSE\left(\frac{1}{n}+\frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right)}} \sim t_{n-2}$$

Linear Models: Matrix Notation

- The general form of a linear model is given by:

$$\tilde{Y} = \mathbf{X}\tilde{\beta} + \tilde{\varepsilon}$$

- Where $\tilde{Y}$ is the $N \times 1$ vector of observed responses
  $\mathbf{X}$ is the $N \times p$ design matrix of fixed constants
  $\tilde{\beta}$ is the $p \times 1$ vector of fixed, but unknown parameters
  $\tilde{\varepsilon}$ is the $N \times 1$ vector of (unobserved) errors

- Class practice: use matrix formulation to write the SLR model.

LS Estimation (Matrix)

- Goal is to estimate: $\mathrm{E}(\tilde{Y}) = \mathbf{X}\tilde{\beta}$

- An estimate $\hat{\beta}$ is the LS estimate of β if and only if:

$$\left(Y - \mathbf{X}\hat{\beta}\right)'\left(Y - \mathbf{X}\hat{\beta}\right) = \min(Y - \mathbf{X}\beta)'(Y - \mathbf{X}\beta)$$

- Multiplying the right term and taking the derivatives generates the normal equations:

$$\mathbf{X}'\mathbf{X}\,\beta = \mathbf{X}'\mathrm{Y}$$

- Keep in mind that <u>$\beta, \hat{\beta}$ and Y are vectors</u> and <u>X is the design matrix</u>!

LS Estimation (Matrix)

- If $\mathbf{X}'\mathbf{X}$ is non-singular, then the *unique* least squares estimates are:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathrm{Y} = (\hat{\beta}_0, \hat{\beta}_1)'$$

- If $(\mathbf{X}'\mathbf{X})^{-1}$ exists then:

  (1) The LS estimate is unbiased: $\mathrm{E}(\hat{\beta}) = \beta$.

  (2) The variance-covariance matrix of LS estimates is given by:

$$\mathrm{cov}(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

- In-class derivation: show (1) and (2).