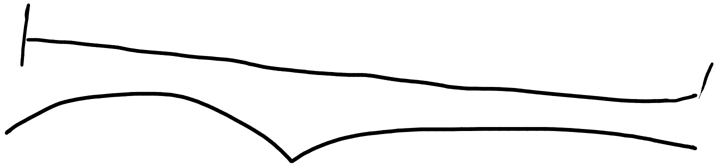
Linear Regression



1. wing span X
2. Mass Y

$$y = mx + b$$

m = slope

b = intercept

$m > 0$ positive

$m = 0$ no

$m < 0$ negative

XY

$XY \sim F(\theta)$

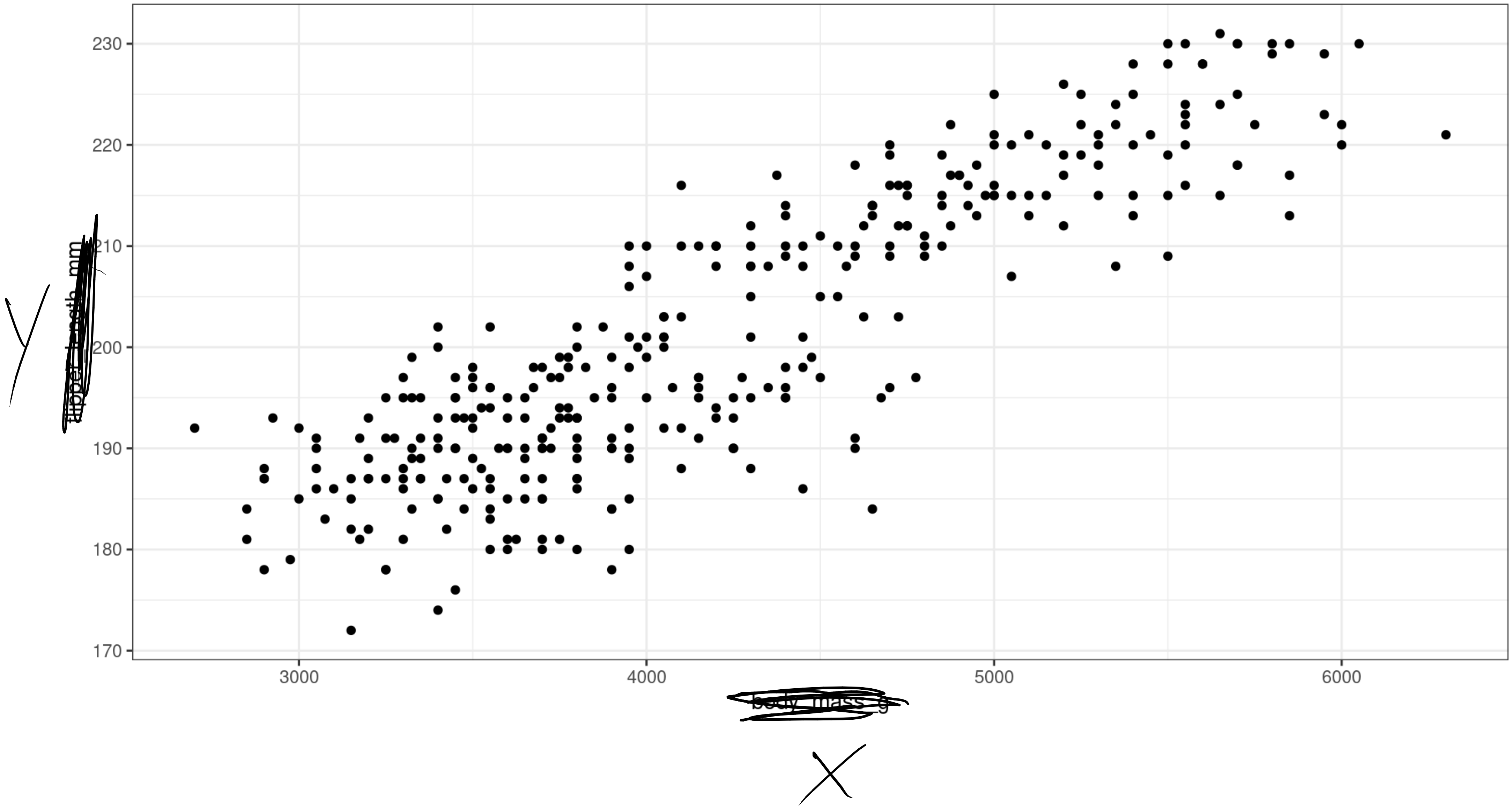
$\{X_i, Y_i\}_{i=1}^{n}$

Learning Outcomes

- Scatter Plot
- Linear Regression
- Ordinary Least Squares
- Unbiasedness

Scatter Plot

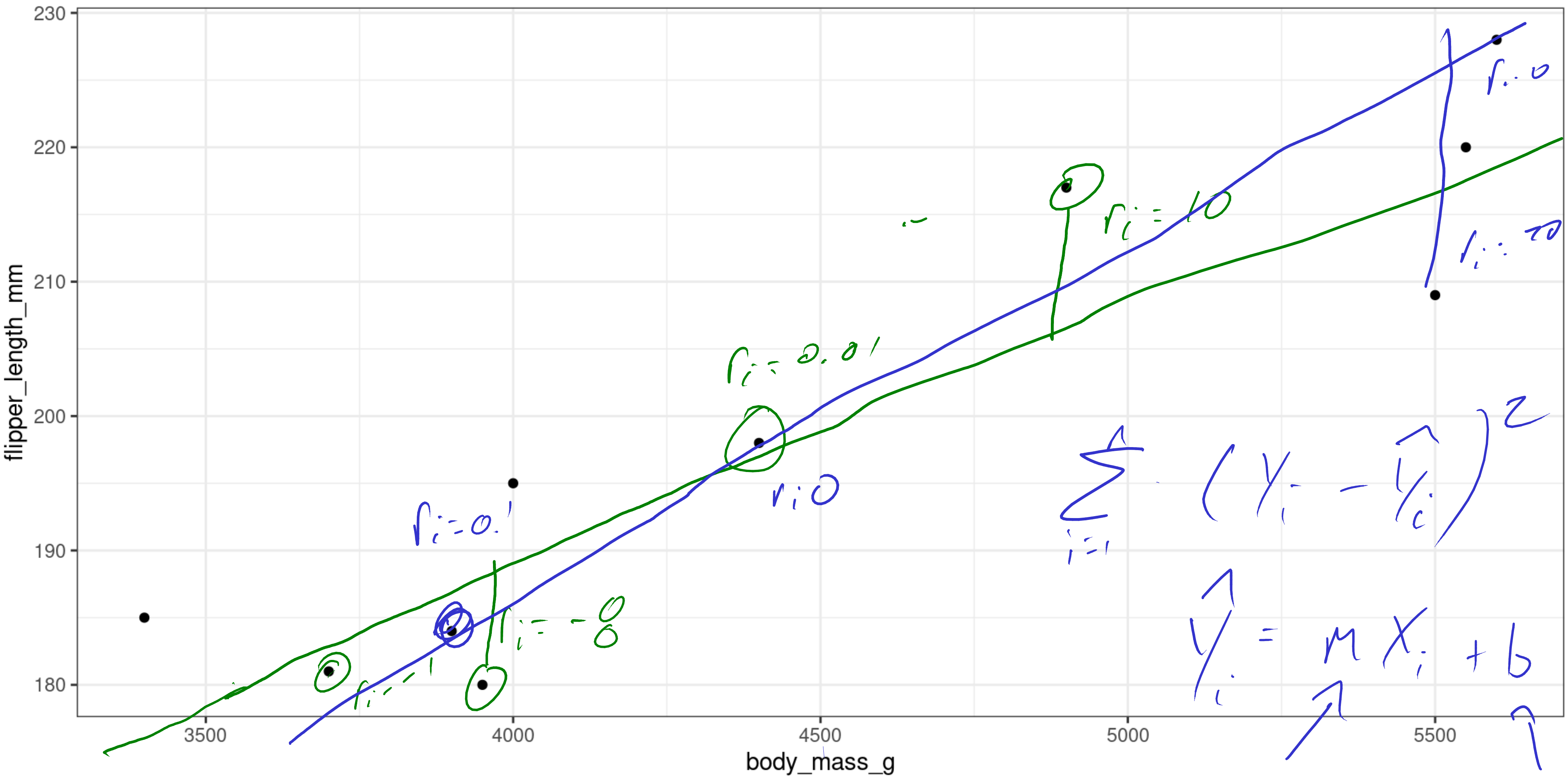
Scatter Plot



Scatter Plot

$r_i = Y_i - \hat{Y}_i$

$\hat{Y}_i \leftarrow$ predicted value result from line



$$\sum_{i=1} (Y_i - \hat{Y}_i)^2$$

$$\hat{Y}_i = m X_i + b$$

Linear Regression

Linear Regression

Linear regression is used to model the association between a set of predictor variables (x's) and an outcome variable (y). Linear regression will fit a line that best describes the data points.

Simple Linear Regression

Simple linear regression will model the association between one predictor variable and an outcome:

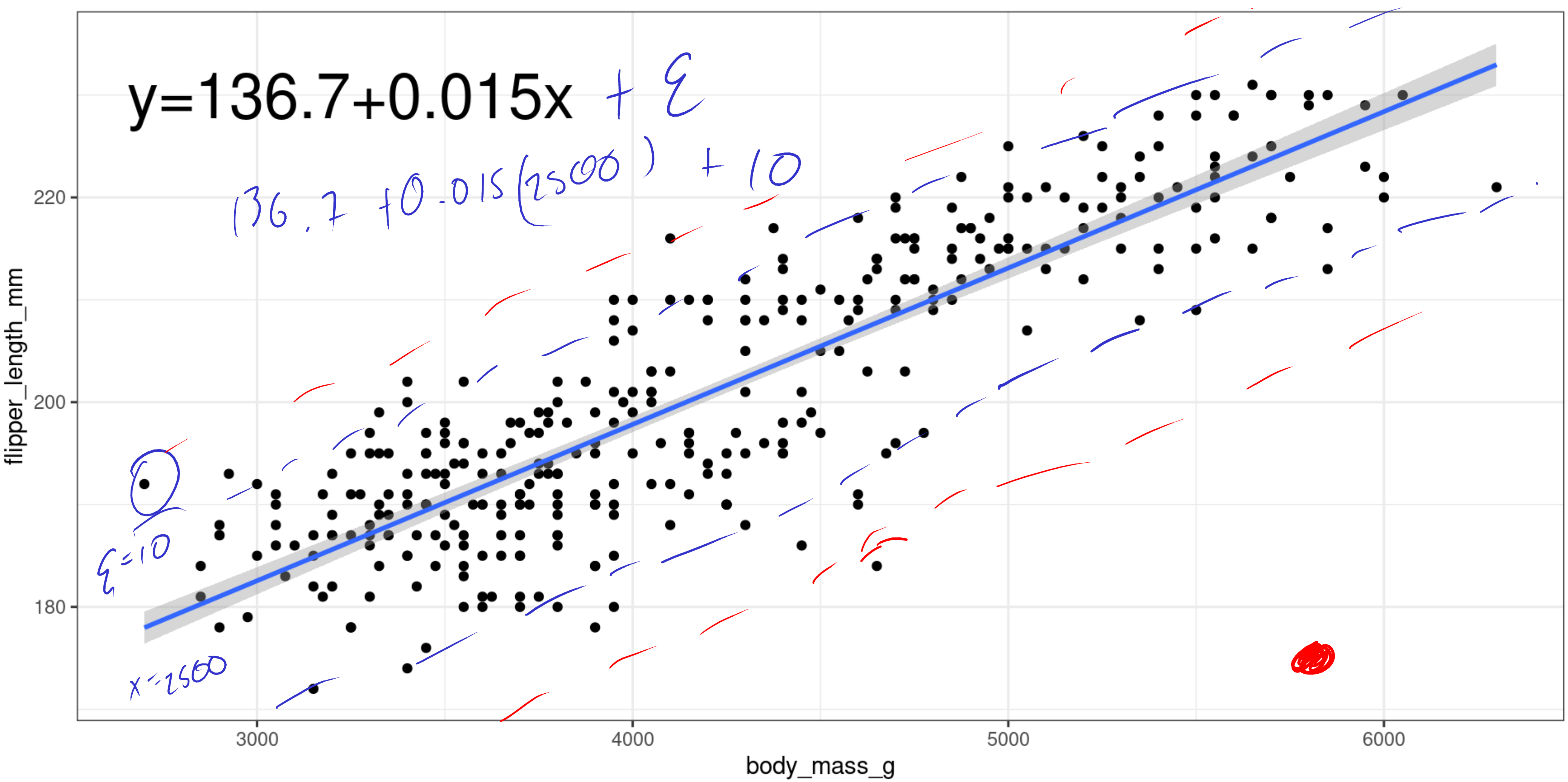
$$Y = \underbrace{\overset{b}{\beta_0} + \overset{m}{\beta_1} X}_{\text{systematic}} + \epsilon$$

random

- β_0: Intercept term
- β_1: Slope term
- $\epsilon \sim N(0, \sigma^2)$

Fitting a Line

$N(0, \sigma^2)$



Interpretation

$$\hat{y} = 136.73 + 0.015x$$

$x = 0$ $\hat{y} = 136.73$

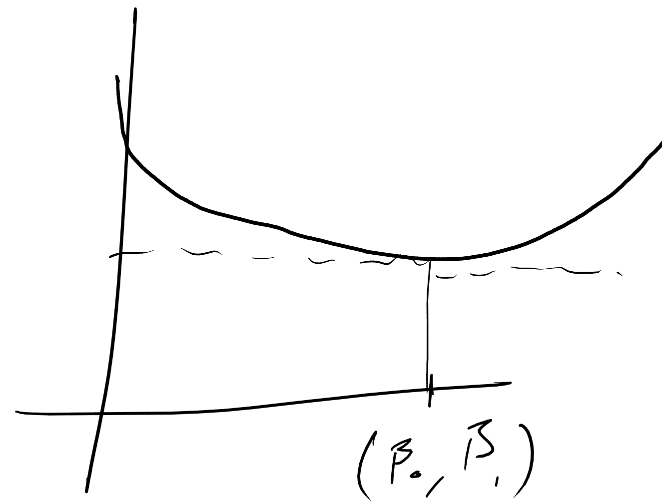
As x increase by 1 unit, y will (+/) increase by an average of 0.015 units.

Ordinary Least Squares

Ordinary Least Squares

For a data pair $(X_i, Y_i)_{i=1}^n$, the ordinary least squares estimator will find the estimates of $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimize the following function:

$$\ell = \sum_{i=1}^{n} \{y_i - (\beta_0 + \beta_1 x_i)\}^2$$



Estimating β's

$$\frac{d\ell}{d\beta_0} = \qquad = 0$$

$$\frac{d\ell}{d\beta_1} = \qquad = 0$$

Estimating β_0

$$\sum_{i=1}^{n} \left(Y_i - (\beta_0 + \beta_1 X_i)\right)^2$$

$$\hat{\beta}_0$$

$$\sum 2\left(Y_i - (\beta_0 + \beta_1 X_i)\right)(-1) = 0$$

$$-2 \sum \left(Y_i - (\beta_0 + \beta_1 X_i)\right) = 0$$

$$\sum \left(Y_i - (\beta_0 + \beta_1 X_i)\right) = 0$$

$$\bar{Y} = \frac{\sum Y_i}{n}$$

$$\sum \left(Y_i - \beta_0 - \beta_1 X_i\right) = 0$$

$$\bar{X} = \frac{\sum X_i}{n}$$

$$\sum Y_i - \sum \beta_0 - \sum \beta_1 X_i = 0$$

$$n\bar{Y} - n\beta_0 - \beta_1 n\bar{X} = 0$$

$$\beta_0 n = n\bar{y} - \beta_1 n\bar{x}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Estimating β_0

$$\sum \left(y_i - (\beta_0 + \beta_1 x_i) \right)^2$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

$$\frac{d\ell}{d\beta_1} = -2 \sum x_i \left(y_i - (\beta_0 + \beta_1 x_i) \right) = 0$$

$$\sum x_i \left(y_i - (\bar{y} - \beta_1 \bar{x} + \beta_1 x_i) \right) = 0$$

$$\sum x_i \left(y_i - \bar{y} + \beta_1 \bar{x} - \beta_1 x_i \right) = 0$$

$$\sum \left(x_i y_i - x_i \bar{y} + \underbrace{\beta_1 x_i \bar{x} - \beta_1 x_i^2} \right) = 0$$

$$\sum x_i y_i - \bar{y} \sum x_i - \beta_1 \sum \left(-x_i \bar{x} + x_i^2 \right) = 0$$

$$\sum x_i y_i - \bar{y} \sum x_i = \beta_1 \sum (x_i^2 - x_i \bar{x})$$

$$\sum (x_i y_i - x_i \bar{y}) = \beta_1 \sum x_i (x_i - \bar{x})$$

$$\sum x_i (y_i - \bar{y}) = \beta_1 \sum x_i (x_i - \bar{x})$$

$$\sum \bar{x} (y_i - \bar{y}) = 0$$

$$\sum \bar{x} (x_i - \bar{x}) = 0$$

$$\hat{\beta}_1 = \frac{\sum x_i (y_i - \bar{y}) + \sum \bar{x} (y_i - \bar{y})}{\sum x_i (x_i - \bar{x}) + \sum \bar{x} (x_i - \bar{x})}$$

$$\bar{x} \sum (y_i - \bar{y}) = 0$$

$$\bar{x} \left(\sum y_i - n\bar{y} \right) = 0$$

$$\bar{x} (n\bar{y} - n\bar{y}) = 0$$

$$\beta_1 = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$$

Estimates

$Y = \beta_0 + \beta_1 X + \varepsilon$

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

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Unbiasedness of β's

Unbiasedness of β's

Both β_0 and β_1 are unbiased estimators.

$E(\hat{\beta}_1) = \beta_1$ $\qquad$ $E(\hat{\beta}_0) = \beta_0$

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

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

$E(\beta_1)$