# Lecture Notes I

# - Simple Linear Regression -

June 25, 2016

## 1 Basic Structure of Simple Linear Regression Model

### 1.1 The description of the Population

**Definition 1.** *The conditional expectation function $E(Y|X)$ is given by*

$$E(Y|X) = \int_{-\infty}^{\infty} yp(y|x)dy \tag{1}$$

**Theorem 1.** *We can describe the relationship between the dependent variable $Y$ and the independent variable $X$ as*

$$Y = E(Y|X) + \epsilon \tag{2}$$

*where*
*(1) $\epsilon$ is mean- independent [1] of $X$, i.e, $E(\epsilon|X) = 0$*
*(2) $\epsilon$ is uncorrelated with any function of $X$*

*Proof.* For (1),

$$E(\epsilon|X) = E(Y - E(Y|X)|X) = E(Y|X) - E(E(Y|X)|X) = E(Y|X) - E(Y|X) = 0 \tag{4}$$

For (2),

$$Cov(h(X), \epsilon) = E(h(X)\epsilon) = E(E(h(X)\epsilon|X)) = E(h(X)E(\epsilon|X)) = 0 \tag{5}$$

□

[1] A random variable $Y$ is said to be mean independent of $X$ if and only if

$$E(Y|X) = E(Y) \tag{3}$$

**Theorem 2.** *Let $m(X)$ be any function of $X$. Then $CEF$ is the best predictor of $Y$ given $X$*

$$E(Y|X) = argmin_{m(X)}\ E(Y - m(X))^2 \tag{6}$$

*Proof.*

$$(Y - m(X))^2 = [(Y - E(Y|X)) + (E(Y|X) - m(X))]^2 \tag{7}$$

$$= (Y - E(Y|X))^2 + 2(Y - E(Y|X))(E(Y|X) - m(X)) + (E(Y|X) - m(X))^2$$

Since $E[2(Y - E(Y|X))(E(Y|X) - m(X))] = 0$ by Theorem 1, $E(Y - m(X))^2$ is minimized when $m(X) = E(Y|X)$ □

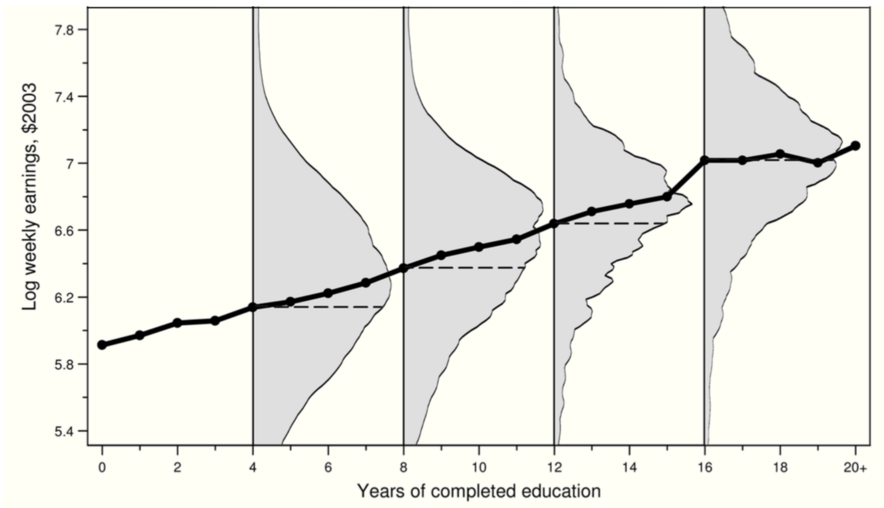


Figure 1: Conditional Expectation (Function)

## 1.2 Gauss-Markov Assumptions

$$\text{A.1} \quad y = \beta_0 + \beta_1 x + \epsilon \tag{8}$$

$$\text{A.2} \quad \sum_{i=1}^{n} (x_i - \bar{x})^2 \neq 0 \tag{9}$$

$$\text{A.3} \quad E(\epsilon|x) = 0 \tag{10}$$

$$\text{A.4} \quad Var(\epsilon|x) = \sigma^2, \quad Cov(\epsilon_i, \epsilon_j) = 0 \quad \forall\ i, j \tag{11}$$

$A.1$ and $A.3$ together imply that the $CEF$ is linear , i.e , $E(Y|X) = \beta_0 + \beta_1 X$

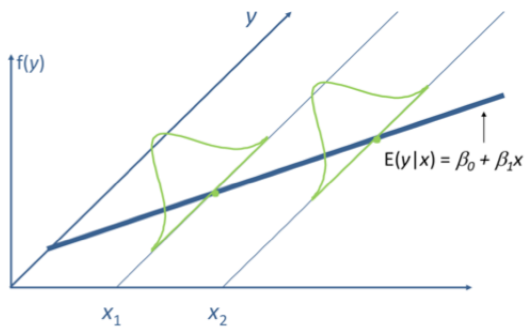


Figure 2: Gauss-Markov Assumptions

## 1.3 Sample

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \qquad\qquad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \tag{12}$$

## 1.4 Scatter Plot

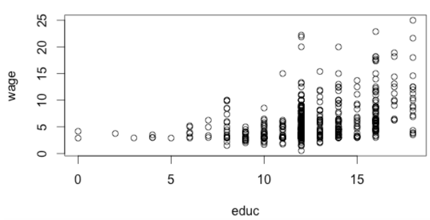


Figure 3: Scatter Plot

## 2 Point Estimation : Ordinary Least Square Method

### 2.1 Basic Idea

**Definition 2.** *Residual vector*
*Let's assume that $\hat{\beta}$ is an estimator of $\beta$.*
*Then residual vector $e$ is defined to be*

$$e = y - \hat{y} = y - X\hat{\beta} \tag{13}$$

*where*

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \qquad X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \qquad \hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} \tag{14}$$

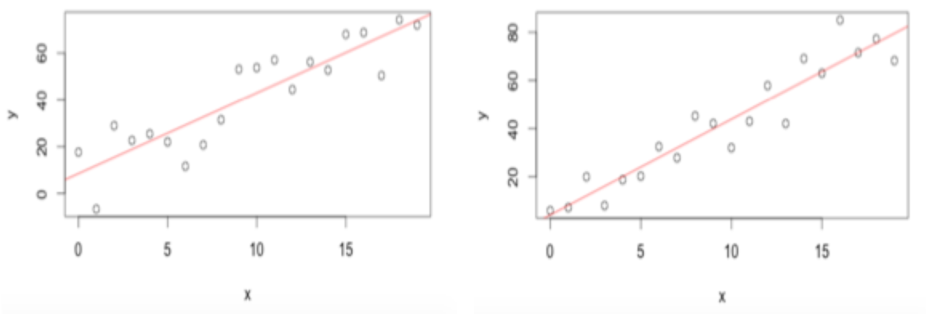


Figure 4: Ordinary Least Square Method

**Definition 3.** *Ordinary Least Square Estimator*
*Ordinary Least Square Estimator $\hat{\beta}$ is the function from $(y, X)$ to $R^2$ that minimize the sum of squares of residuals.*

$$\hat{\beta} = argmin \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \tag{15}$$

### 2.2 The Derivation of OLS Estimator $\hat{\beta}_{OLS}$

$$min \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \tag{16}$$

F.O.C

$$\frac{\partial}{\partial \hat{\beta_0}} \sum_{i=1}^{n} e_i^2 = -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = -2\Sigma e_i = 0 \tag{17}$$

$$\frac{\partial}{\partial \hat{\beta_1}} \sum_{i=1}^{n} e_i^2 = -2 \sum_{i=1}^{n} x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = -2\Sigma x_i e_i = 0 \tag{18}$$

By rearranging,

$$\overline{y} = \hat{\beta}_0 + \hat{\beta}_1 \overline{x} \tag{19}$$

$$\sum_{i=1}^{n} x_i y_i = \hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 \tag{20}$$

These equations are called the $Normal\ Equations$

From the Normal equations, OLS Estimator $\hat{\beta}_{OLS}$ can be deduced as follows

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})y_i}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \tag{21}$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} \tag{22}$$

### 2.3 The Derivation of OLS Estimator $\hat{\beta}_{OLS}$ in matrix form

$$min\ e'e = (y - X\hat{\beta})'(y - X\hat{\beta}) \tag{23}$$

$$\frac{\partial}{\partial \hat{\beta}}\ e'e = -2X'y + 2X'X\hat{\beta} = 0 \tag{24}$$

By rearranging,

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'y \tag{25}$$

### 2.4 Geometric Interpretation of $OLS$ Estimator $\hat{\beta}_{OLS}$

To find the OLS estimator is equal to find the projection of $y$ on $X$

$$\hat{\beta} = argmin \; \|e\|^2 = argmin\|e\| \tag{26}$$

where $\|e\|$ is the norm of residual vector $e$

$$X'e = X'(y - X'\hat{\beta}) = 0 \tag{27}$$

By rearranging ,

$$X'X\hat{\beta} = X'y \tag{28}$$

$$\hat{\beta} = (X'X)^{-1}X'y \tag{29}$$

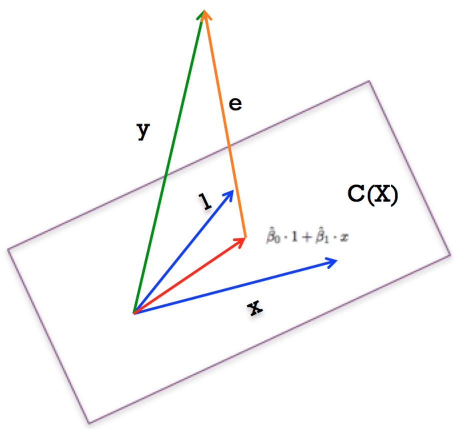


Figure 5: Projection of $y$ on $C(X)$

### 2.5 The OLS Estimator as Sample Analog Estimator(or $MOM$ Estimator)

In the population

$$E(\epsilon) = E(y - X\beta) = 0 \tag{30}$$

$$E(x\epsilon) = E[x(y - X\beta)] = 0 \tag{31}$$

The sample counterpart of the above equations are

$$\tfrac{1}{n}\Sigma e_i = \tfrac{1}{n}\Sigma(y_i - (X\hat{\beta})_i) = 0 \tag{32}$$
$$\tfrac{1}{n}\Sigma x_i e_i = \tfrac{1}{n}\Sigma x_i(y_i - (X\hat{\beta})_i) = 0 \tag{33}$$

or in matrix form

$$X'e = X'(y - X\hat{\beta}) = 0 \tag{34}$$

Thus, $OLS$ estimator is a kind of sample analogue estimator(or $MOM$ estimator)

# 3 The Properties of OLS Estimators, $\hat{\beta}_{OLS}$

## 3.1 Unbiasedness

**Theorem 3.** *Under the Assumptions $A.1$ through $A.3$,*

$$E(\hat{\beta}) = \beta \tag{35}$$

*Proof.*

$$\hat{\beta}_1 = \frac{\sum(x_i - \overline{x})y_i}{\sum(x_i - \overline{x})^2} = \frac{\sum(x_i - \overline{x})(\beta_0 + \beta_1 x_i + \epsilon_i)}{\sum(x_i - \overline{x})^2} = \beta_1 + \frac{\sum(x_i - \overline{x})\epsilon_i}{\sum(x_i - \overline{x})^2} \tag{36}$$

$$E(\hat{\beta}_1) = \beta_1 + \sum \frac{(x_i - \overline{x})}{\sum(x_i - \overline{x})^2} E(\epsilon_i) = \beta_1 \tag{37}$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x} = (\beta_0 + \beta_1\overline{x} + \overline{\epsilon}) - \hat{\beta}_1\overline{x} = \beta_0 + (\beta_1 - \hat{\beta}_1)\overline{x} + \overline{\epsilon} \tag{38}$$

$$E(\hat{\beta}_0) = \beta_0 + (\beta_1 - E(\hat{\beta}_1))\overline{x} + E(\overline{\epsilon}) = \beta_0 \tag{39}$$

□

## 3.2 Variances of the $OLS$ Estimator

**Theorem 4.** *Under the Assumption $A.1$ through $A.4$,*

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum(x_i - \overline{x})^2} \tag{40}$$

*Proof.*

$$\hat{\beta}_1 - \beta_1 = \frac{\sum(x_i - \overline{x})\epsilon_i}{\sum(x_i - \overline{x})^2} \tag{41}$$

$$Var(\hat{\beta}_1) = E(\hat{\beta}_1 - \beta_1)^2 = E(\frac{\sum(x_i - \overline{x})\epsilon_i}{\sum(x_i - \overline{x})^2})^2 = \frac{\sum(x_i - \overline{x})^2 Var(\epsilon_i)}{(\sum(x_i - \overline{x})^2))^2} = \frac{\sigma^2}{\sum(x_i - \overline{x})^2} \tag{42}$$

□

### 3.3 Gauss-Markov Theorem

**Theorem 5.** *Under the Assumption A.1 through A.4,*
*OLS estimators $\hat{\beta}$ are the Best Linear Unbiased Estimators(BLUEs) of $\beta$*

*Proof.* The linear estimator of $\beta_1$ must take the form

$$\hat{\beta}_1 = \sum \omega_i y_i \tag{43}$$

For a linear estimator $\beta_1$ to be unbiased, the following conditions should be satisfied

$$\sum \omega_i = 0 \tag{44}$$

$$\sum \omega_i x_i = 1 \tag{45}$$

The variance of $\beta_1$ is equal to

$$Var(\hat{\beta}_1) = Var(\sum \omega_i y_i) = \sum \omega_i^2 Var(y_i) = \sigma_y^2 \sum \omega_i^2 \tag{46}$$

Now to find the best linear unbiased estimator means to fine $\omega_i, \forall$ to minimize the variance of $\beta_1$ subject to the constraints (44) and (45) That is,

$$min_{\omega_i} \sum \omega_i^2 \ s.t \sum \omega_i = 0, \sum \omega_i x_i = 1 \tag{47}$$

Setting up the Lagrangian function,

$$L = \sum \omega_i^2 - \mu_1 \sum \omega_i - \mu_2(\sum \omega_i \tilde{x_i} - 1) \tag{48}$$

where $\tilde{x_i} = x_i - \overline{x}$ The first conditions are as follows

$$\frac{\partial L}{\partial \omega_i} = 2\omega_i - \mu_1 - \mu_2 \tilde{x_i} = 0, \forall i \tag{49}$$

$$\frac{\partial L}{\partial \mu_1} = \sum \omega_i = 0 \tag{50}$$

$$\frac{\partial L}{\partial \mu_2} = \sum \omega_i \tilde{x_i} - 1 = 0 \tag{51}$$

The equation (49) implies

$$\omega_i = \frac{\mu_1}{2} + \frac{\mu_2 \tilde{x_i}}{2} \tag{52}$$

Substituting (52) into (50) yields

$$\mu_1 = 0 \tag{53}$$

thus

$$\omega_i = \frac{\mu_2 \tilde{x_i}}{2} \tag{54}$$

Substituting (54) into (51) yields

$$\mu_2 = \frac{2}{\sum \tilde{x_i}^2} \tag{55}$$

Finally, Substituting (53) and (55) into (52) yields

$$\omega_i = \frac{\tilde{x_i}}{\sum \tilde{x_i}^2} = \frac{x_i - \overline{x}}{\sum (x_i - \overline{x})^2} \quad \forall i \tag{56}$$

,which implies that $\hat{\beta}_{1OLS}$ is $BLUE$ □

## 4 The Estimator of $\sigma^2, \hat{\sigma}^2$

**Theorem 6.** *Let's define $\hat{\sigma}^2$ as*

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-2} \tag{57}$$

*Then,under the Assumption A.1 through A.4,*

$$E(\hat{\sigma}^2) = \sigma^2 \tag{58}$$

*Proof.*

$$\begin{aligned} e_i &= y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) = (\beta_0 + \beta_1 x_i + \epsilon_i) - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \\ &= (\beta_0 - \hat{\beta}_0) + (\beta_1 - \hat{\beta}_1)x_i + \epsilon_i \end{aligned} \tag{59}$$

From (16),

$$\begin{aligned} \sum e_i &= \sum[(\beta_0 - \hat{\beta}_0) + (\beta_1 - \hat{\beta}_1)x_i + \epsilon_i] \\ &= n(\beta_0 - \hat{\beta}_0) + (\beta_1 - \hat{\beta}_1)\sum x_i + \sum \epsilon_i = 0 \end{aligned} \tag{60}$$

Dividing both sides of (60) by $n$,then

$$\overline{\epsilon} + (\beta_0 - \hat{\beta}_0) + (\beta_1 - \hat{\beta}_1)\overline{x} = 0 \tag{61}$$

From (58) and (60),

$$e_i = (\beta_1 - \hat{\beta}_1)(x_i - \overline{x}) + (\epsilon_i - \overline{\epsilon}) \tag{62}$$

Thus,

$$\begin{aligned} \sum e_i^2 &= \sum(\epsilon_i - \overline{\epsilon})^2 + (\beta_1 - \hat{\beta}_1)^2 \sum(x_i - \overline{x})^2 + 2(\beta_1 - \hat{\beta}_1)\sum(\epsilon_i - \overline{\epsilon})(x_i - \overline{x}) \\ &= \sum(\epsilon_i - \overline{\epsilon})^2 - (\beta_1 - \hat{\beta}_1)^2 \sum(x_i - \bar{x})^2 \end{aligned} \tag{63}$$

And

$$E(\sum e_i^2) = (n-1)\sigma^2 - \sigma^2 = (n-2)\sigma^2 \tag{64}$$

$$\therefore E(\hat{\sigma}^2) = \sigma^2 \tag{65}$$

□

**Definition 4.** *The standard error of $\hat{\beta}_1$ is defined as*

$$se(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{\sum(x_i - \overline{x})^2}} \tag{66}$$

## 5 Goodness-of-Fit

### 5.1 The decomposition of $SST$

**Definition 5.** *Total Sum of Squares($SST$), Explained Sum of Squares($SSE$) and Residual Sum of Squares($SSR$) are defined as*

$$SST \equiv \sum (y_i - \overline{y})^2 \tag{67}$$
$$SSE \equiv \sum (\hat{y}_i - \overline{y})^2 \tag{68}$$
$$SSR \equiv \sum (y_i - \hat{y}_i)^2 \tag{69}$$

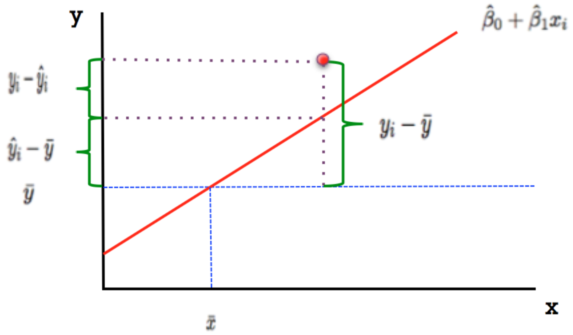


Figure 6: SST,SSE and SSR

**Theorem 7.** *$SST$ can be decomposed into $SSE$ and $SSR$, that is,*

$$SST = SSE + SSR \tag{70}$$

*Proof.*

$$\begin{aligned} \sum (y_i - \overline{y})^2 &= \sum [(y_i - \hat{y}_i) + (\hat{y}_i - \overline{y})]^2 \\ &= \sum (y_i - \hat{y}_i)^2 + \sum (\hat{y}_i - \overline{y})^2 + 2 \sum (y_i - \hat{y}_i)(\hat{y}_i - \overline{y}) \end{aligned} \tag{71}$$

Since

$$\sum (y_i - \hat{y}_i)(\hat{y}_i - \overline{y}) = 0, \tag{72}$$

$$SST = SSE + SSR \tag{73}$$

□

### 5.2 The coefficient of Determination $R^2$

**Definition 6.** *The coefficient of determination $R^2$ is defined as*

$$R^2 \equiv \frac{SSE}{SST} = 1 - \frac{SSR}{SST} \tag{74}$$

$R^2$ is an estimator for the population R-squared given by [2]

$$1 - \frac{\sigma_\epsilon^2}{\sigma_y^2} \tag{78}$$

$R^2$ is an measure of how well the least squares equation

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \tag{79}$$

performs as a predictor of $y$

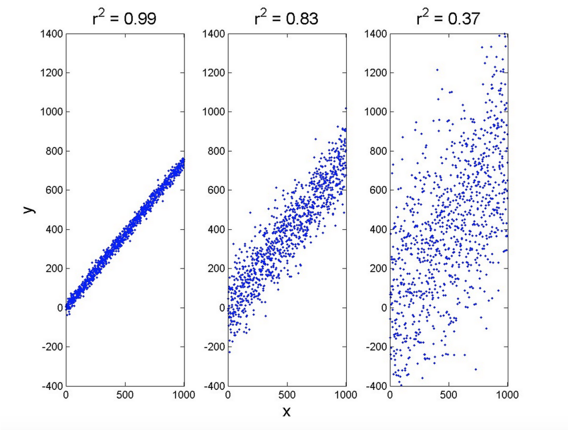


Figure 7: The coefficient of Determination

[2]The population R-squared can be deduced as follows.
By decomposition of variance,

$$Var(Y) = E_X(Var(Y|X)) + Var_X(E(Y|X)) \tag{75}$$

Dividing both sides of the equation by $Var(Y)$ and rearranging, we get

$$\frac{Var_X(E(Y|X))}{Var(Y)} = 1 - \frac{E_X(Var(Y|X))}{Var(Y)} \tag{76}$$

Since $Var(Y|X) = \sigma_\epsilon^2$,

$$\frac{Var_X(E(Y|X))}{Var(Y)} = 1 - \frac{\sigma_\epsilon^2}{\sigma_y^2} \tag{77}$$

### 5.3 The relationship between the coefficient of determination $R^2$ and the sample correlation coefficient $\rho_{xy}$ in the simple regression model

**Theorem 8.** *In the simple regression model, the coefficient of determination $R^2$ is equal to the square of sample correlation coefficient between $y$ and $x$. That is,*

$$R^2 = \rho_{xy}^2 \tag{80}$$

*Proof.*

$$R^2 = \frac{\sum(\hat{y}_i - \overline{y})^2}{\sum(y_i - \overline{y})^2} \tag{81}$$

Since

$$\hat{y}_i - \overline{y} = (\hat{\beta}_0 + \hat{\beta}_1 x_i) - (\hat{\beta}_0 + \hat{\beta}_1 \overline{x}) = \hat{\beta}_1(x_i - \overline{x}) \tag{82}$$

and

$$\hat{\beta}_1 = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sum(x_i - \overline{x})^2} \tag{83}$$

we can rewrite $R^2$ as

$$R^2 = \hat{\beta}_1^2 \frac{\sum(x_i - \overline{x})^2}{\sum(y_i - \overline{y})^2} = \frac{(\sum(x_i - \overline{x})(y_i - \overline{y}))^2}{\sum(x_i - \overline{x})^2 \sum(y_i - \overline{y})^2} \tag{84}$$

Thus

$$R^2 = \rho_{xy}^2 \tag{85}$$

□

## 6 Regression through the Origin

Suppose

$$y_i = \beta_1 x_i + \epsilon_i \tag{86}$$

In this case, the *OLS* estimator $\hat{\beta_1}*$ of $\beta_1$ is[3]

$$\hat{\beta_1}* = \frac{\sum x_i y_i}{\sum x_i^2} \tag{89}$$

[3] $\hat{\beta_1}*$ can be derived by the following procedure

$$min \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \tag{87}$$

$$\frac{d}{d\hat{\beta}_1}(\sum e_i^2) = -2\sum(y_i - \hat{\beta}_1 x_i) = 0 \tag{88}$$

Unless $\overline{x} = 0$, $\hat{\beta_1}*$ is different from $\hat{\beta_1}$

$$\hat{\beta_1}* = \frac{\sum x_i y_i}{\sum x_i^2} \neq \frac{\sum (x_i - \overline{x}) y_i}{\sum (x_i - \overline{x})^2} = \hat{\beta_1} \tag{90}$$

If $\beta_0 \neq 0$, then $\hat{\beta_1}*$ is not unbiased.

$$\hat{\beta_1}* = \frac{\sum x_i(\beta_0 + \beta_1 x_1 + \epsilon_i)}{\sum x_i^2} = \frac{\sum x_i}{\sum x_i^2}\beta_0 + \beta_1 + \frac{\sum x_i}{\sum x_i^2}\epsilon_i \tag{91}$$

$$E(\hat{\beta_1}*) = \frac{\sum x_i}{\sum x_i^2}\beta_0 + \beta_1 + \frac{\sum x_i}{\sum x_i^2}E(\epsilon_i) = \frac{\sum x_i}{\sum x_i^2}\beta_0 + \beta_1 \neq \beta_1 \tag{92}$$