

## Regression Analysis (I)

Kutner's Applied Linear Statistical Models (5/E)

### Chapter 1: Linear Regression with One Predictor Variable

Thursday 09:10-12:00, 商館 260205

**Han-Ming Wu**

Department of Statistics, National Chengchi University

<http://www.hmwu.idv.tw>

## Overview

1. Regression analysis (迴歸分析) is a \_\_\_\_\_ that utilizes the relation between two or more \_\_\_\_\_ so that a \_\_\_\_\_ or \_\_\_\_\_ variable can be predicted from the other, or others.
2. Examples: general form of a regression model \_\_\_\_\_ :
  - (a)  $Y$ : the sales of a product,  $X$ : the amount of advertising expenditures (支出).
  - (b)  $Y$ : the performance of an employee on a job,  $X$ : a battery of aptitude tests (能力傾向成套測驗, 性向測驗).
  - (c)  $Y$ : the size of the vocabulary of a child,  $X_1$ : age of the child,  $X_2$ : amount of education of the parents.
  - (d)  $Y$ : the length of hospital stay of a surgical patient,  $X_1$ : the time in the hospital,  $X_2$ : the severity of the operation.
3. In this chapter, we consider the basic ideas of regression analysis and discuss the \_\_\_\_\_ of regression models containing a single predictor variable.

## 1.1 Relations between Variables

### Functional Relation between Two Variables

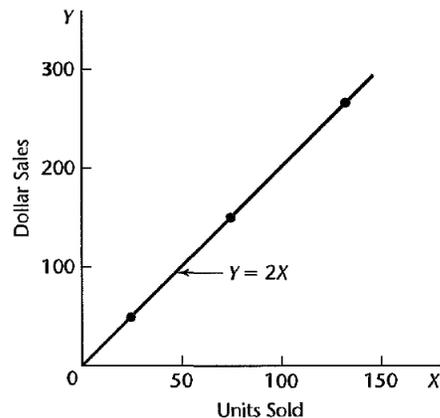
1. A \_\_\_\_\_ relation between two variables is expressed by a \_\_\_\_\_ formula. If  $X$  denotes the \_\_\_\_\_ variable and  $Y$  the \_\_\_\_\_ variable, a functional relation is of the form:

\_\_\_\_\_

2. **Example:**  $Y$ : dollar sales of a product sold at a fixed price,  $X$ : the number of units sold. If the selling price is \$2 per unit, the relation is expressed by the equation:

\_\_\_\_\_.

**FIGURE 1.1**  
Example of  
Functional  
Relation.

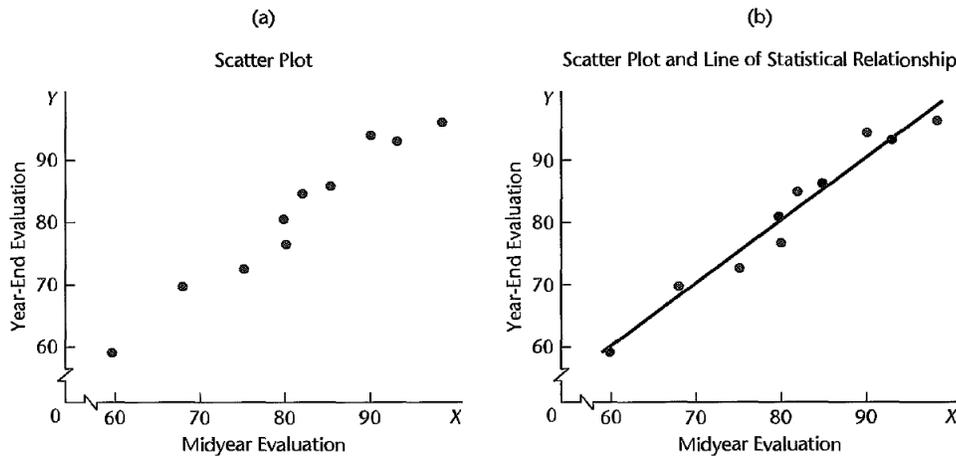


### Statistical Relation between Two Variables

1. In general, the observations for a statistical relation do not \_\_\_\_\_ the curve of relationship.
2. **Example 1:** Performance evaluations
- Performance evaluations for 10 employees were obtained at midyear ( $X$ ) and at year-end ( $Y$ ).
  - Figure 1.2a: the \_\_\_\_\_ the midyear evaluation, the \_\_\_\_\_ tends to be the year-end evaluation.
  - Figure 1.2b: a \_\_\_\_\_ that describes the statistical relation between midyear and year-end evaluations.

- (d) Note: that most of the points do not fall directly on the line of statistical relationship. This \_\_\_\_\_ around the line represents \_\_\_\_\_ in year-end evaluations that is not associated with midyear performance evaluation and that is usually considered to be of a \_\_\_\_\_.

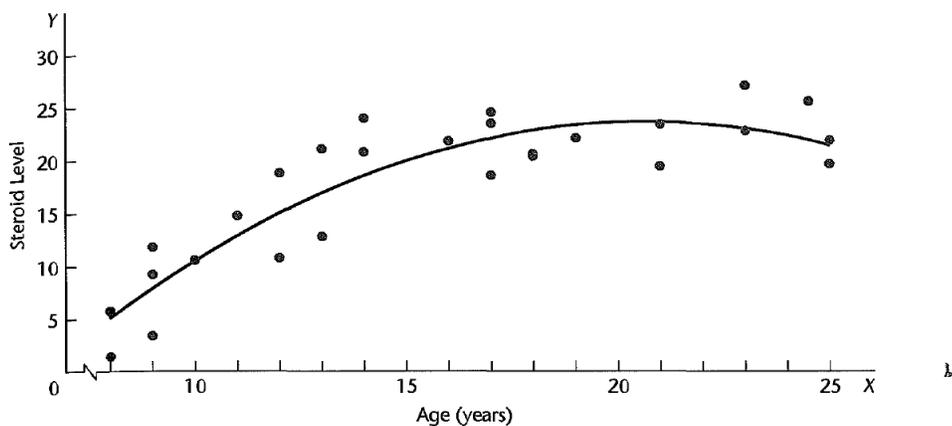
**FIGURE 1.2 Statistical Relation between Midyear Performance Evaluation and Year-End Evaluation.**



3. Example 2:

- (a) The data on age and level of a steroid (類固醇) in plasma (血漿) for 27 healthy females between 8 and 25 years old. (Figure 1.3)
- (b) The data strongly suggest that the statistical relationship is \_\_\_\_\_ (not linear).
- (c) As age \_\_\_\_\_, steroid level \_\_\_\_\_ up to a point and then begins to \_\_\_\_\_.

**FIGURE 1.3 Curvilinear Statistical Relation between Age and Steroid Level in Healthy Females Aged 8 to 25.**



## 1.2 Regression Models and Their Uses

### Historical Origins

1. Regression analysis was first developed by \_\_\_\_\_ in the latter part of the \_\_\_\_\_.
2. Galton had studied the relation between \_\_\_\_\_ and noted that the heights of children of both tall and short parents appeared to \_\_\_\_\_ (回復) or \_\_\_\_\_ (回歸) to the \_\_\_\_\_.
3. He considered this tendency to be a regression to \_\_\_\_\_.
4. Galton developed a mathematical description of this \_\_\_\_\_, the precursor of today's regression models.
5. The term regression persists to this day to describe \_\_\_\_\_.



行銷資料科學: 小時了了, 大未必佳 迴歸均值的有趣現象:

<https://medium.com/marketingdatascience/d5f8e5e73163>.



均值迴歸 (regression toward the mean) 現象: 當一個特性的極端傾向發生時, 會有返回這項特性的平均值 (regression toward mediocrity)。



例子: 身高較高的父母, 其子女的平均身高, 要低於他們父母的平均身高, 不會長得更高; 相對的, 身高比較矮的父母, 其子女的平均身高, 要高於他們父母的平均身高, 不會變得更矮。

### Basic Concepts

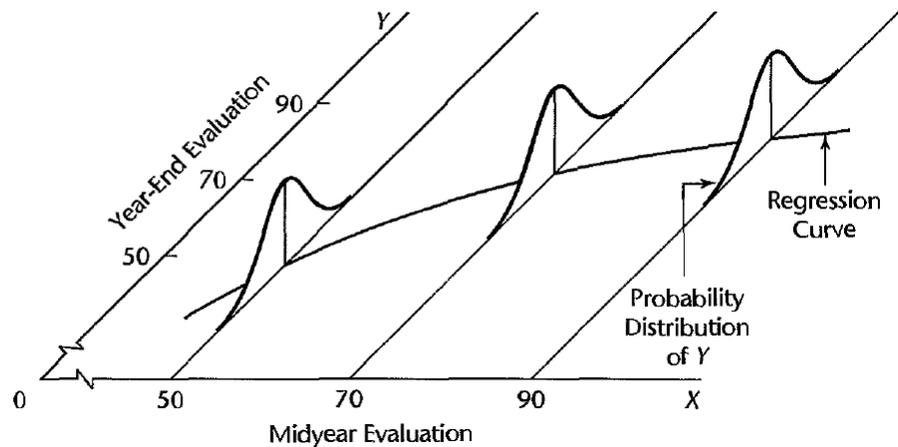
1. A regression model is:
  - (a) A tendency of the \_\_\_\_\_ variable  $Y$  to vary with the \_\_\_\_\_ variable  $X$  in a \_\_\_\_\_ fashion.
  - (b) A scattering of points around the \_\_\_\_\_ of statistical relationship.
2. Assumptions for a regression model:
  - (a) There is a \_\_\_\_\_ (機率分佈) of  $Y$  for each level of  $X$ .

- (b) The \_\_\_\_\_ of these probability distributions vary in some systematic fashion with \_\_\_\_\_. (\_\_\_\_\_)

3. **Example:** Performance evaluation (Figure 1.2)

- (a) The year-end evaluation  $Y$  is treated in a regression model as a \_\_\_\_\_. For each level of midyear performance evaluation \_\_\_\_\_, there is postulated a \_\_\_\_\_.

**FIGURE 1.4**  
Pictorial Representation of Regression Model.



- (b) Figure 1.4: shows probability distributions of  $Y$  for midyear evaluation levels at  $X = 50$ ,  $X = 70$  and  $X = 90$ . Note that the \_\_\_\_\_ of the probability distributions have a systematic relation to the level of  $X$ .
- (c) This systematic relationship is called the \_\_\_\_\_. The graph of the regression function is called the \_\_\_\_\_.
- (d) The regression curve, which describes the relation between \_\_\_\_\_ and \_\_\_\_\_, is the counterpart to the general tendency of  $Y$  to vary with  $X$  systematically in a statistical relation.

## Construction of Regression Models

### 1. Selection of Predictor Variables:

- (a) Choosing a \_\_\_\_\_ of explanatory or \_\_\_\_\_ variables that is "good" in some sense for the purposes of the analysis.

- (b) Other considerations: the \_\_\_\_\_ of the variable; the degree to which observations on the variable can be obtained more \_\_\_\_\_ than on competing variables; and the degree to which the variable can be \_\_\_\_\_.

## 2. Functional Form of Regression Relation:

- (a) The functional form of the regression relation is \_\_\_\_\_ and must be decided upon \_\_\_\_\_ once the data have been collected.
- (b) The \_\_\_\_\_ or \_\_\_\_\_ regression functions are often used as satisfactory first approximations to regression functions of unknown nature.

## 3. Scope of Model:

- (a) In formulating a regression model, we usually need to \_\_\_\_\_ of the model to some interval or region of values of the predictor variable(s).
- (b) **Example:** a company studying the effect of price on sales volume investigated six price levels, ranging from \$4.95 to \$6.95. Here, the scope of the model is limited to price levels ranging from near \$5 to near \$7. The shape of the regression function substantially outside this range would be in serious doubt because the investigation provided no evidence as to the nature of the statistical relation below \$4.95 or above \$6.95.

## Uses of Regression Analysis

1. Regression analysis serves three major purposes: (1) \_\_\_\_\_, (2) \_\_\_\_\_, and (3) \_\_\_\_\_.
2. The several purposes of regression analysis frequently \_\_\_\_\_ in practice.

## Regression and Causality (因果關係)

1. The existence of a statistical relation between the response variable  $Y$  and the explanatory or predictor variable  $X$  \_\_\_\_\_ in any way that  $Y$  depends \_\_\_\_\_ on  $X$ .

2. No matter how strong is the statistical relation between  $X$  and  $Y$ , no \_\_\_\_\_ pattern is necessarily implied by the regression model.
3. **Example:** data on size of vocabulary ( $X$ ) and writing speed ( $Y$ ) for a sample of young children aged 5-10 will show a positive regression relation. This relation does not imply, however, that an increase in vocabulary causes a faster writing speed. Here, other explanatory variables, such as age of the child and amount of education, affect both the vocabulary ( $X$ ) and the writing speed ( $Y$ ). Older children have a larger vocabulary and a faster writing speed.
4. Regression analysis by itself provides \_\_\_\_\_ about causal patterns and must be supplemented by \_\_\_\_\_ to obtain insights about causal relations.

## Use of Computers

1. Regression analysis often entails lengthy and tedious calculations, computers are usually utilized to perform the necessary calculations.
2. Almost every statistics package for computers contains a regression component: BMDP, MINITAB, \_\_\_\_\_, \_\_\_\_\_, SYSTAT, JMP, S-Plus, MATLAB, and \_\_\_\_\_.

## 1.3 Simple Linear Regression Model with Distribution of Error Terms Unspecified

### Formal Statement of Model

1. A simple linear regression model:

$$\text{_____} \quad (1.1)$$

where:

- (a)  $Y_i$ : the value of the \_\_\_\_\_ variable in the \_\_\_\_\_.
- (b)  $\beta_0$  and  $\beta_1$ : \_\_\_\_\_ to be estimated.

- (c)  $X_i$ : the value of the \_\_\_\_\_ variable in the  $i$ th trial
- (d)  $\epsilon_i$ : a \_\_\_\_\_ term with mean \_\_\_\_\_ and variance \_\_\_\_\_.
- (e)  $\epsilon_i$  and  $\epsilon_j$  are \_\_\_\_\_ so that their covariance is zero (i.e., \_\_\_\_\_ for all  $i, j; i \neq j$ )  $i = 1, \dots, n$ .
2. Regression model (1.1) is said to be
- (a) simple: there is \_\_\_\_\_ predictor variable
- (b) linear in the \_\_\_\_\_: no parameter appears as an exponent or is multiplied or divided by another parameter
- (c) linear in the \_\_\_\_\_ variable: because this variable appears only in the first power.
3. A model that is linear in the parameters and in the predictor variable is also called \_\_\_\_\_ model.

### Important Features of Model

1. The response  $Y_i$  in the  $i$ th trial is the sum of two components: (1) the constant term \_\_\_\_\_ and (2) the random term \_\_\_\_\_. Hence,  $Y_i$  is a \_\_\_\_\_.
2. Since  $E(\epsilon_i) = 0$ , it follows that:

$$E(Y_i) = \underline{\hspace{2cm}} = \underline{\hspace{2cm}} = \underline{\hspace{2cm}}.$$

Thus, the response  $Y_i$ , when the level of  $X$  in the  $i$ th trial is  $X_i$ , comes from a probability distribution whose mean is:

$$\underline{\hspace{2cm}}.$$

The regression function for model (1.1) is:

$$\underline{\hspace{2cm}}$$

since the regression function relates the means of the probability distributions of  $Y$  for given  $X$  to the level of  $X$ .

3. The response  $Y_i$  in the  $i$ th trial \_\_\_\_\_ of the value of the regression function ( \_\_\_\_\_ ) by the error term amount \_\_\_\_\_.
4. The error terms  $\epsilon_i$  are assumed to have constant variance \_\_\_\_\_. It therefore follows that the responses  $Y_i$  have the same constant variance:

$$\sigma^2(Y_i) = \sigma^2$$

Thus, regression model (1.1) assumes that the probability distributions of  $Y$  have the same variance \_\_\_\_\_, regardless of the level of the predictor variable  $X$ .

5. Since the error terms  $\epsilon_i$  and  $\epsilon_j$  are assumed to be uncorrelated, so are the responses \_\_\_\_\_.
6. **Summary:** regression model \_\_\_\_\_ implies that the responses  $Y_i$  come from probability distributions whose means are \_\_\_\_\_ and whose variances are \_\_\_\_\_, the same for all levels of  $X$ . Further, any two responses  $Y_i$  and  $Y_j$  are \_\_\_\_\_.

7. **Example:** Electrical distributor (Figure 1.6)

A consultant for an electrical distributor is studying the relationship between the number of bids ( \_\_\_\_\_ ) requested by construction contractors for basic lighting equipment during a week and the number of hours ( \_\_\_\_\_ ) required to prepare the bids.

- (a) Suppose that regression model (1.1) is:

$$Y_i = 9.5 + 2.1X_i + \epsilon_i$$

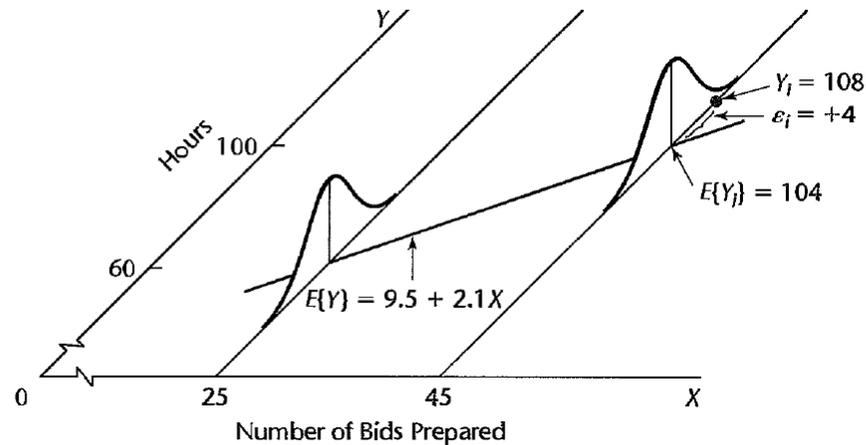
- (b) The regression function is:
- \_\_\_\_\_

- (c) Suppose that in the  $i$ th week,  $X_i = 45$  bids are prepared and the actual number of hours required is  $Y_i = 108$ . We have

$$E(Y_i) = \underline{\hspace{2cm}} \quad \text{and} \quad \epsilon_i = \underline{\hspace{2cm}}$$

- (d) The error term  $\epsilon_i$  is simply the \_\_\_\_\_ of  $Y_i$  from its mean value  $E(Y_i)$ .

**FIGURE 1.6**  
**Illustration of**  
**Simple Linear**  
**Regression**  
**Model (1.1).**



## Meaning of Regression Parameters

1. The parameters  $\beta_0$  and  $\beta_1$ , in regression model (1.1) are called \_\_\_\_\_.

(a) The parameter  $\beta_0$  is the  $Y$  \_\_\_\_\_ of the regression line.  $\beta_1$ , is the \_\_\_\_\_ of the regression line.

(b)  $\beta_1$  indicates the \_\_\_\_\_ in the mean of the probability distribution of  $Y$  per unit increase in  $X$ .

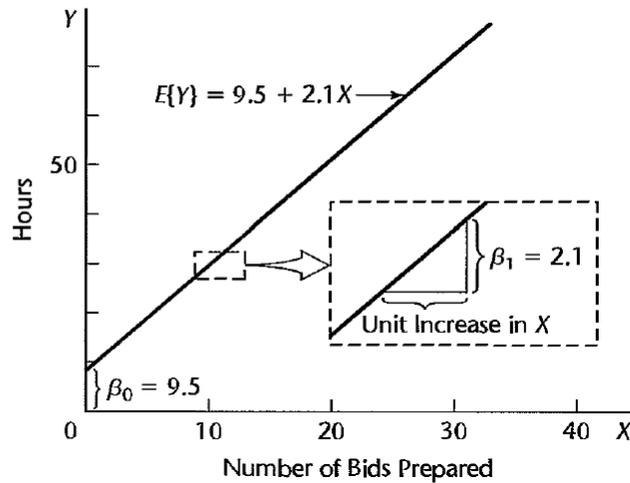
(c) When the scope of the model includes \_\_\_\_\_,  $\beta_0$  gives the mean of the probability distribution of  $Y$  at  $X = 0$ . When the scope of the model does not cover  $X = 0$ ,  $\beta_0$  \_\_\_\_\_ as a separate term in the regression model.

2. **Example:** Electrical distributor (Figure 1.7)

(a) The regression function:  $E(Y) = 9.5 + 2.1X$ . The slope  $\beta_1 = 2.1$  indicates that the preparation of \_\_\_\_\_ bid in a week leads to an \_\_\_\_\_ in the \_\_\_\_\_ of the probability distribution of  $Y$  of 2.1 hours.

(b) The intercept  $\beta_0 = 9.5$  indicates the value of the regression function at \_\_\_\_\_. Since the linear regression model was formulated to apply to weeks where the number of bids prepared ranges from \_\_\_\_\_,  $\beta_0 = 9.5$  does not have any intrinsic meaning of its own here.

**FIGURE 1.7**  
**Meaning of**  
**Parameters of**  
**Simple Linear**  
**Regression**  
**Model (1.1).**



### Alternative Versions of Regression Model

1. Let \_\_\_\_\_ be a constant identically equal to \_\_\_\_\_. Then, we can write (1.1) as follows:

$$\text{_____} \quad \text{where } X_0 \equiv 1$$

This version of the model associates an  $X$  variable with each regression coefficient.

2. An alternative modification is to use for the predictor variable the \_\_\_\_\_ rather than  $X_i$ :

$$\begin{aligned}
 Y_i &= \text{_____} \\
 &= \text{_____} \\
 &= \text{_____},
 \end{aligned}$$

where

$$\beta_0^* = \text{_____}$$

## 1.4 Data for Regression Analysis\*

## 1.5 Overview of Steps in Regression Analysis\*

## 1.6 Estimation of Regression Function

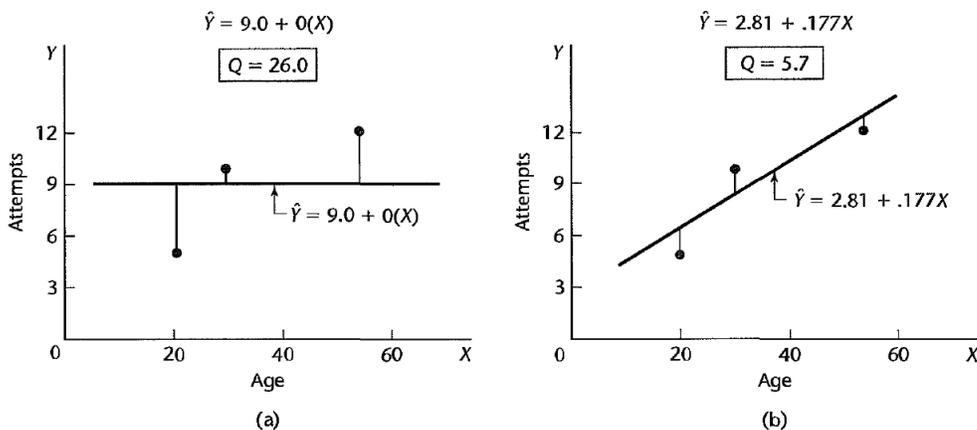
### Method of least Squares

1. For the observations \_\_\_\_\_ for each case, the method of least squares considers the sum of the  $n$  squared deviation of  $Y_i$  from its expected value  $E(Y_i)$ :

$$Q = \text{_____} \quad (1.8)$$

2. According to the method of least squares, the estimators of  $\beta_0$  and  $\beta_1$  are those values  $b_0$  and  $b_1$  respectively, that \_\_\_\_\_ for the given sample observations  $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ .

**FIGURE 1.9 Illustration of Least Squares Criterion  $Q$  for Fit of a Regression Line—Persistence Study Example.**



3. **Example:** (Figure 1.9)

- (a) Figure 1.9a:  $Y = 9.0 + 0 \cdot X$ . This regression line is not a good fit. The sum of the squared deviations for the three cases is:

$$Q = (5 - 9.0)^2 + (12 - 9.0)^2 + (10 - 9.0)^2 = 26.0$$

- (b) Figure 1.9b:  $Y = 2.81 + 0.177X$  (the least squares regression line). The criterion  $Q$  is much reduced:

$$Q = (5 - 6.35)^2 + (12 - 12.55)^2 + (10 - 8.12)^2 = 5.7$$

Thus, a better fit of the regression line to the data corresponds to a smaller sum  $Q$ .

#### 4. Least Squares Estimators:

- (a) For given sample observations  $(X_i, Y_i)$ , the quantity  $Q$  in (1.8) is a function of  $\beta_0$  and  $\beta_1$ . The values of  $\beta_0$  and  $\beta_1$ , that minimize  $Q$  can be derived by differentiating (1.8) with respect to  $\beta_0$  and  $\beta_1$ :

$$\frac{\partial Q}{\partial \beta_0} = \underline{\hspace{10em}}$$

$$\frac{\partial Q}{\partial \beta_1} = \underline{\hspace{10em}}$$

- (b) Set these partial derivatives equal to zero, using  $b_0$  and  $b_1$  (or                     ) to denote the particular values of  $\beta_0$  and  $\beta_1$ , that minimize  $Q$ :

$$-2 \sum (Y_i - b_0 - b_1 X_i) = 0 \Rightarrow \underline{\hspace{10em}}$$

$$-2 \sum X_i (Y_i - b_0 - b_1 X_i) = 0 \Rightarrow \underline{\hspace{10em}}.$$

- (c) Normal equations:

$$\begin{aligned} & \underline{\hspace{10em}} \\ & \underline{\hspace{10em}}, \\ & b_0 \text{ and } b_1 \text{ are called point estimators of } \beta_0 \text{ and } \beta_1, \text{ respectively.} \end{aligned}$$

- (d) The normal equations can be solved simultaneously for  $b_0$  and  $b_1$ :

$$b_0 = \underline{\hspace{10em}} = \underline{\hspace{10em}}$$

$$b_1 = \underline{\hspace{10em}}$$

where  $\bar{X}$  and  $\bar{Y}$  are the means of the  $X_i$  and the  $Y_i$  observations, respectively.

## 5. Properties of Least Squares Estimators:

- (a) **Gauss-Markov theorem:** Under the conditions of regression model (1.1), the least squares estimators  $b_0$  and  $b_1$  in (1.10) are \_\_\_\_\_ and have \_\_\_\_\_ among all unbiased linear estimators.

$$\text{_____ and _____,}$$

so that neither estimator tends to overestimate or underestimate systematically.

- (b) The theorem states that the estimators  $b_0$  and  $b_1$  are \_\_\_\_\_ (i.e., their sampling distributions are \_\_\_\_\_) than any other estimators belonging to the class of unbiased estimators that are linear functions of the observations  $Y_1, \dots, Y_n$ .
- (c) The estimators  $b_0$  and  $b_1$  are such linear functions of the  $Y_i$ .

$$b_1 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}$$

This expression is equal to:

$$b_1 = \text{_____} = \text{_____}$$

where:

$$k_i = \text{_____}$$

Since the  $k_i$  are known constants (because the  $X_i$  are known constants),  $b_1$  is a linear combination of the  $Y_i$  and hence is a linear estimator.

- (d) In the same fashion, it can be shown that  $b_0$  is a linear estimator.

6. **Example:** The Toluca Company Manufactures Refrigeration Equipment

In the past, one of the replacement parts has been produced periodically in lots of varying sizes. When a cost improvement program was undertaken, company officials wished to determine the optimum lot size for producing this part. The production of this part involves setting up the production process and machining and assembly operations. One key input for the model to ascertain the optimum lot size was the relationship between lot size and labor hours required to produce the

lot. To determine this relationship, data on lot size and work hours for 25 recent production runs were utilized. The production conditions were stable during the six-month period in which the 25 runs were made and were expected to continue to be the same during the next three years, the planning period for which the cost improvement program was being conducted.

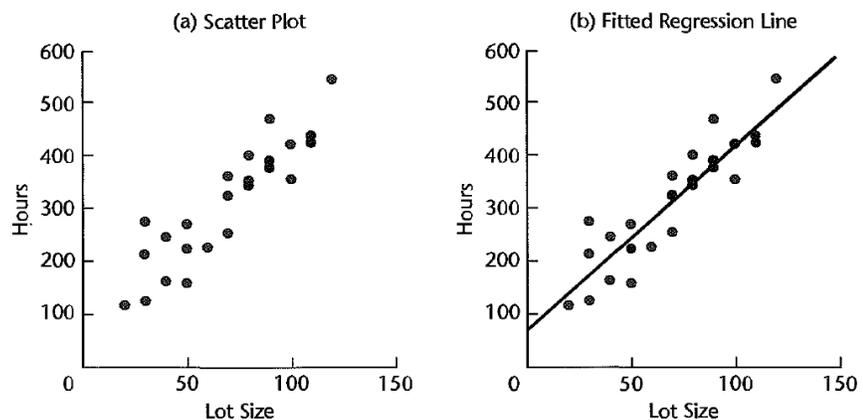
- (a) (Table 1.1) All lot sizes are multiples of 10, a result of company policy to facilitate the administration of the parts production.

**TABLE 1.1 Data on Lot Size and Work Hours and Needed Calculations for Least Squares Estimates—Toluca Company Example.**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Run	Lot Size	Work Hours					
$i$	$X_i$	$Y_i$	$X_i - \bar{X}$	$Y_i - \bar{Y}$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$
1	80	399	10	86.72	867.2	100	7,520.4
2	30	121	-40	-191.28	7,651.2	1,600	36,588.0
3	50	221	-20	-91.28	1,825.6	400	8,332.0
...	...	...	...	...	...	...	...
23	40	244	-30	-68.28	2,048.4	900	4,662.2
24	80	342	10	29.72	297.2	100	883.3
25	70	323	0	10.72	0.0	0	114.9
Total	1,750	7,807	0	0	70,690	19,800	307,203
Mean	70.0	312.28					

- (b) (Figure 1.10a) shows a SYSTAT scatter plot of the data. The scatter plot indicates that the relationship between \_\_\_\_\_ and \_\_\_\_\_ is reasonably \_\_\_\_\_. We also see that no observations on work hours are \_\_\_\_\_, with reference to the relationship between lot size and work hours.

**FIGURE 1.10 SYSTAT Scatter Plot and Fitted Regression Line—Toluca Company Example.**



(c) Calculate the least squares estimates:

$$b_1 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2} = \frac{70690}{19800} = 3.5702$$

$$b_0 = \bar{Y} - b_1\bar{X} = 312.28 - 3.5702(70.0) = 62.37$$

(d) We estimate that the \_\_\_\_\_ number of work hours \_\_\_\_\_ for each additional unit produced in the lot. This estimate applies to the range of lot sizes (from about \_\_\_\_\_ to about \_\_\_\_\_) in the data from which the estimates were derived.

**FIGURE 1.11** The regression equation is

$$Y = 62.4 + 3.57 X$$

Portion of  
MINITAB

Regression

Output—

Toluca

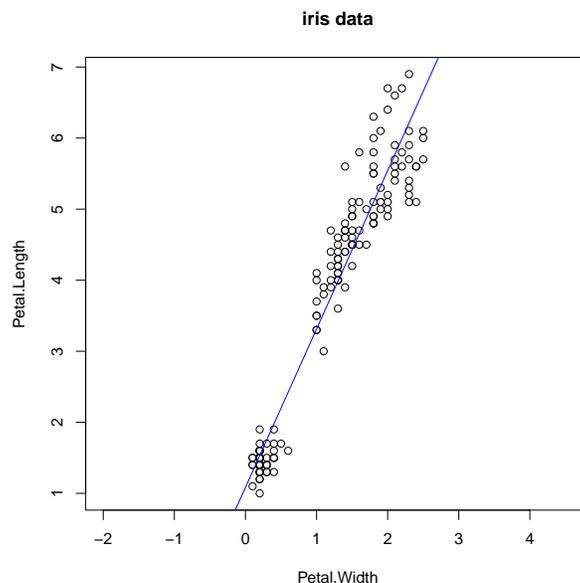
Company

Example.

Predictor	Coef	Stdev	t-ratio	p
Constant	62.37	26.18	2.38	0.026
X	3.5702	0.3470	10.29	0.000

s = 48.82      R-sq = 82.2%      R-sq(adj) = 81.4%

😊 R code example:



```

> head(iris)
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1           5.1           3.5           1.4           0.2  setosa
2           4.9           3.0           1.4           0.2  setosa
3           4.7           3.2           1.3           0.2  setosa
4           4.6           3.1           1.5           0.2  setosa
5           5.0           3.6           1.4           0.2  setosa
6           5.4           3.9           1.7           0.4  setosa
> str(iris)
'data.frame':  150 obs. of  5 variables:
 $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
 $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
 $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
 $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
 $ Species      : Factor w/ 3 levels "setosa","versicolor",...: 1 1 1 1 1 1 1 1 1 1 ...
> attach(iris)
> plot(Petal.Width, Petal.Length, main = "iris data", asp = 1)
> iris.lm <- lm(Petal.Length ~ Petal.Width)
> summary(iris.lm)

Call:
lm(formula = Petal.Length ~ Petal.Width)

Residuals:
    Min       1Q   Median       3Q      Max
-1.33542 -0.30347 -0.02955  0.25776  1.39453

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.08356    0.07297   14.85  <2e-16 ***
Petal.Width  2.22994    0.05140   43.39  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4782 on 148 degrees of freedom
Multiple R-squared:  0.9271,    Adjusted R-squared:  0.9266
F-statistic: 1882 on 1 and 148 DF,  p-value: < 2.2e-16

> abline(iris.lm, col = "blue")

```

## Point Estimation of Mean Response

### 1. Estimated Regression Function

- (a) Given sample estimators  $b_0$  and  $b_1$  of the parameters in the regression function:

\_\_\_\_\_

we estimate the regression function as follows:

\_\_\_\_\_

where  $\hat{Y}$  (read \_\_\_\_\_) is the value of the estimated regression function at the level  $X$  of the predictor variable.

- (b) We call a value of the response variable a \_\_\_\_\_ and  $E(Y)$  the \_\_\_\_\_.
- (c) The mean response stands for the mean of the probability distribution of  $Y$  corresponding to the level  $X$  of the predictor variable.
- (d)  $\hat{Y}$  then is a point estimator of the mean response when the level of the predictor variable is  $X$ .
- (e) An extension of the Gauss-Markov theorem:  $\hat{Y}$  is an \_\_\_\_\_ estimator of  $E(Y)$ , with \_\_\_\_\_ in the class of unbiased linear estimators.
- (f) For the cases in the study, we will call  $\hat{Y}_i$ :

$$\hat{Y}_i = \text{_____}, \quad i = 1, \dots, n$$

the \_\_\_\_\_ for the  $i$ th case. Thus, the fitted value  $\hat{Y}_i$  is to be viewed in distinction to the \_\_\_\_\_.

### 2. Example: The Toluca Company Example

- (a) (Figure 1.10b) The estimated regression function:

$$\hat{Y} = 62.37 + 3.5702X$$

It appears to be a good description of the \_\_\_\_\_ between lot size and work hours.

- (b) Suppose that we estimate the mean number of work hours (mean response) required when the lot size is  $X = 65$  units:

$$\hat{Y} = \underline{\hspace{10em}} \text{ hours}$$

- (c) Interpretation: if many lots of 65 units are produced under the conditions of the 25 runs on which the estimated regression function is based, the mean labor time for these lots is about 294 hours.

- (d) **NOTE** Of course, the labor time for anyone lot of size 65 is likely to fall above or below the mean response because of inherent variability in the production system, as represented by the error term in the model.)

- (e) (Table 1.2) The fitted value for the first case  $X_1 = 80$  is:

$$\hat{Y}_1 = \underline{\hspace{10em}} \text{ hours}$$

**TABLE 1.2**  
Fitted Values,  
Residuals, and  
Squared  
Residuals—  
Toluca  
Company  
Example.

	(1)	(2)	(3)	(4)	(5)
	Lot	Work	Estimated		Squared
Run	Size	Hours	Mean	Residual	Residual
$i$	$X_i$	$Y_i$	Response	$Y_i - \hat{Y}_i = e_i$	$(Y_i - \hat{Y}_i)^2 = e_i^2$
			$\hat{Y}_i$		
1	80	399	347.98	51.02	2,603.0
2	30	121	169.47	-48.47	2,349.3
3	50	221	240.88	-19.88	395.2
...	...	...	...	...	...
23	40	244	205.17	38.83	1,507.8
24	80	342	347.98	-5.98	35.8
25	70	323	312.28	10.72	114.9
<b>Total</b>	<b>1,750</b>	<b>7,807</b>	<b>7,807</b>	<b>0</b>	<b>54,825</b>

☺ *R code example:*

```

> predict(iris.lm, list(Petal.Width = c(0.2, 0.4)))
      1      2
1.529546 1.975534
> data.frame(iris.lm$fitted.values, iris.lm$residuals)
      iris.lm.fitted.values iris.lm.residuals
1             1.529546      -0.129546132
2             1.529546      -0.129546132
3             1.529546      -0.229546132
...
8             1.529546      -0.029546132
9             1.529546      -0.129546132
10            1.306552       0.193447918
...

```

### 3. Alternative Model

(a) When the alternative regression model (1.6) is to be utilized:

\_\_\_\_\_ ,

the least squares estimator  $b_1$  of  $\beta_1$  \_\_\_\_\_ as before.

(b) The least squares estimator of  $\beta_0^* = \beta_0 + \beta_1 \bar{X}$  becomes

$$b_0^* = \underline{\hspace{4cm}}$$

Hence, the estimated regression function for alternative model (1.6) is:

\_\_\_\_\_

### Residuals (残差)

1. The  $i$ th residual is the difference between the \_\_\_\_\_ and the corresponding \_\_\_\_\_. This residual is denoted by \_\_\_\_\_ :

$$e_i = \underline{\hspace{2cm}}$$

2. For regression model (1.1), the residual  $e_i$  becomes:

$$e_i = \underline{\hspace{4cm}}$$

3. (Figure 1.12) The magnitude of a residual is represented by the \_\_\_\_\_ of the  $Y_i$  observation from the corresponding point on the estimated regression function (i.e., from the corresponding fitted value  $\hat{Y}_i$ ).

**NOTE** We need to distinguish between the model error term value \_\_\_\_\_ and the residual \_\_\_\_\_. The former involves the vertical deviation of  $Y_i$  from the unknown true regression line and hence is \_\_\_\_\_. On the other hand, the residual is the vertical deviation of  $Y_i$  from the fitted value  $\hat{Y}_i$  on the estimated regression line, and it is \_\_\_\_\_.

4. Residuals are highly useful for studying whether a given regression model is \_\_\_\_\_ for the data at hand.

## Properties of Fitted Regression Line

1. The sum of the residuals is zero:

$$\sum e_i = \sum (Y_i - b_0 - b_1 X_i) = \sum Y_i - nb_0 - b_1 \sum X_i$$

**NOTE** Rounding errors may, of course, be present in any particular case, resulting in a sum of the residuals that does not equal zero exactly.

2. The sum of the squared residuals, \_\_\_\_\_, is a minimum. This was the requirement to be satisfied in deriving the least squares estimators of the regression parameters.
3. The sum of the observed values  $Y_i$  equals the sum of the fitted values  $\hat{Y}_i$ :

4. The sum of the weighted residuals is zero when the residual in the  $i$ th trial is weighted by the level of the predictor variable in the  $i$ th trial:

5. The sum of the weighted residuals is zero when the residual in the  $i$ th trial is weighted by the fitted value of the response variable for the  $i$ th trial:

6. The regression line always goes through the point \_\_\_\_\_.

$$\hat{Y} = \text{_____} = \bar{Y}$$

## 1.7 Estimation of Error Terms Variance $\sigma^2$

### Point Estimator of $\sigma^2$

1. The variance  $\sigma^2$  of the \_\_\_\_\_ in regression model (1.1) needs to be estimated to obtain an indication of the \_\_\_\_\_ of the probability distributions of  $Y$ . A variety of \_\_\_\_\_ (推論) concerning the regression function and the prediction of  $Y$  require an estimate of  $\sigma^2$ .

2. **Single Population:** The estimator of the variance  $\sigma^2$  is the sample variance  $s^2$ :

$$s^2 = \text{_____}$$

which is an \_\_\_\_\_ estimator of the variance  $\sigma^2$  of an infinite population. The sample variance is often called a \_\_\_\_\_, because a sum of squares has been divided by the appropriate number of \_\_\_\_\_.

### 3. Regression Model

(a) We need to calculate a \_\_\_\_\_, but must recognize that the  $Y_i$  now come from \_\_\_\_\_ probability distributions with \_\_\_\_\_ means that depend upon the level  $X_i$ . The deviations are the \_\_\_\_\_:

\_\_\_\_\_ and the appropriate sum of squares, denoted by \_\_\_\_\_, is:

$$\text{SSE} = \text{_____}$$

where SSE stands for \_\_\_\_\_ or \_\_\_\_\_.

- (b) The sum of squares SSE has \_\_\_\_\_ degrees of freedom associated with it. Two degrees of freedom are lost because both \_\_\_\_\_ had to be estimated in obtaining the estimated means  $\hat{Y}_i$ . Hence, the appropriate \_\_\_\_\_, denoted by MSE or  $s^2$ , is:

$$s^2 = \text{MSE} = \frac{\text{SSE}}{\text{df}}$$

where MSE stands for \_\_\_\_\_ or \_\_\_\_\_.

- (c) It can be shown that MSE is an \_\_\_\_\_ estimator of  $\sigma^2$  for regression model (1.1): \_\_\_\_\_.

4. **Example:** The Toluca Company Example

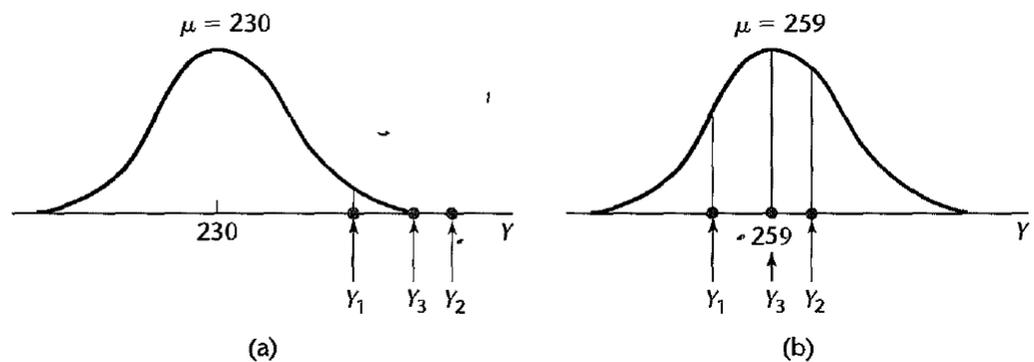
- (a) (Table 1.2) we obtain:  $SSE = 54,825$  and

$$s^2 = \text{MSE} = \frac{54,825}{23} = 2,384$$

A point estimate of  $\sigma$ , the standard deviation of the probability distribution of  $Y$  for any  $X$ , is  $s = \sqrt{2,384} = 48.8$  hours.

- (b) Consider again the case where the lot size is  $X = 65$  units. We found earlier that the mean of the probability distribution of  $Y$  for this lot size is estimated to be 294.4 hours. Now, we have the additional information that the standard deviation of this distribution is estimated to be 48.8 hours.

**FIGURE 1.13**  
Densities for Sample Observations for Two Possible Values of  $\mu$ :  $Y_1 = 250$ ,  $Y_2 = 265$ ,  $Y_3 = 259$ .



## 1.8 Normal Error Regression Model

### Model

1. To set up \_\_\_\_\_ and make \_\_\_\_\_, however, we need to make an assumption about the form of the distribution of the error terms  $\epsilon_i$ : they are \_\_\_\_\_.

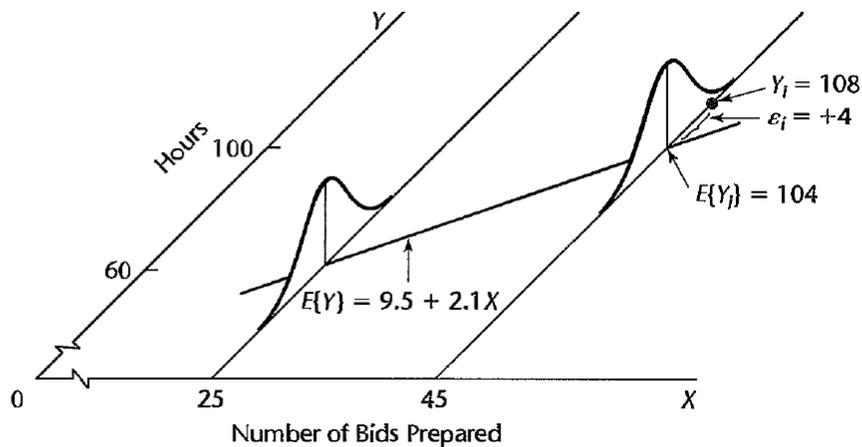
2. The normal error regression model:

$$\text{_____}, \quad i = 1, \dots, n, \quad (1.24)$$

- (a)  $Y_i$ : the \_\_\_\_\_ in the  $i$ th trial.
- (b)  $X_i$ : known constant, the level of the \_\_\_\_\_ variable in the  $i$ th trial.
- (c)  $\beta_0$  and  $\beta_1$ : \_\_\_\_\_ to be estimated.
- (d)  $\epsilon_i$ : independent normally distributed, with mean 0 and variance  $\sigma^2$  (\_\_\_\_\_).

3. (Figure 1.6) Regression model (1.24) implies that the \_\_\_\_\_ are independent normal random variables, with mean \_\_\_\_\_ and variance \_\_\_\_\_.

**FIGURE 1.6**  
**Illustration of**  
**Simple Linear**  
**Regression**  
**Model (1.1).**



4. The normality assumption for the error terms is \_\_\_\_\_ in many situations because

- (a) the error terms frequently represent the \_\_\_\_\_ omitted from the model that \_\_\_\_\_ to some extent and that \_\_\_\_\_ without reference to the variable  $X$ .

- (b) the estimation and testing procedures are based on the \_\_\_\_\_ and are usually only sensitive to large departures from \_\_\_\_\_. Thus, unless the departures from normality are \_\_\_\_\_, particularly with respect to \_\_\_\_\_, the actual confidence coefficients and risks of errors will be close to the levels for \_\_\_\_\_.

## Estimation of Parameters by Method of Maximum likelihood

### 1. Single Population\*

#### 2. Regression Model

- (a) For the normal error regression model (1.24), each  $Y_i$  observation is normally distributed with mean \_\_\_\_\_ and standard deviation \_\_\_\_\_.
- (b) The density of an observation  $Y_i$  for the normal error regression model (1.24) is:

$$f_i = \underline{\hspace{10cm}}$$

- (c) The \_\_\_\_\_ (可能性函数) for  $n$  observations  $Y_1, Y_2, \dots, Y_n$  is the product of the individual densities. Since the variance  $\sigma^2$  of the error terms is usually unknown, the likelihood function is a function of three parameters, \_\_\_\_\_.

$$L(\beta_0, \beta_1, \sigma^2) = \underline{\hspace{10cm}}$$

$$= \underline{\hspace{10cm}}$$

- (d) The values of  $\beta_0$ ,  $\beta_1$ , and  $\sigma^2$  that maximize this likelihood function are the \_\_\_\_\_ (最大概估計量) and are denoted by \_\_\_\_\_, respectively.
- (e) We find the values of  $\beta_0$ ,  $\beta_1$  and  $\sigma^2$  that maximize the logarithm of likelihood function  $\log L$ :

$$\log L = \underline{\hspace{10cm}}.$$



**3. Properties:**

Since the maximum likelihood estimators  $\hat{\beta}_0$  and  $\hat{\beta}_1$ , are the same as the least squares estimators  $b_0$  and  $b_1$  they have the properties of all least squares estimators:

- (a) They are \_\_\_\_\_.
- (b) They have \_\_\_\_\_ among all unbiased linear estimators.
- (c) In addition, the maximum likelihood estimators  $\hat{\beta}_0$  and  $\hat{\beta}_1$ , for the normal error regression model have other desirable properties: \_\_\_\_\_ (A.52), \_\_\_\_\_ (A.53) and the \_\_\_\_\_ estimators (linear or otherwise).

 **Question** ..... (p72)

Assume the normal error regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i.$$

Find the estimation of parameters using method of maximum likelihood.

*sol:*

## TA Class

- **Problems:** 1.6, 1.7, 1.18, 1.20, 1.24
- **Exercises:** 1.32, 1.33, 1.35, 1.36, 1.41
- **Projects:** 1.43