p1

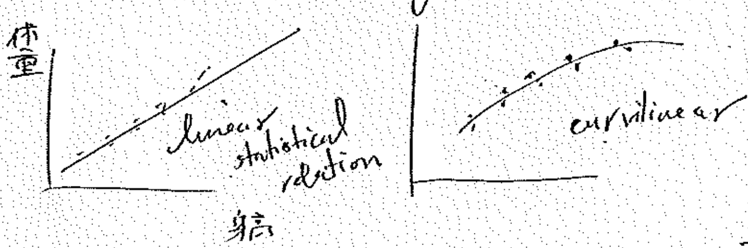
A functional relation between two variables X and Y is exact.

the value of Y is uniquely determined when the value of X is specified.

$$\underset{\text{of electrical motor}}{\overset{\text{rental fee}}{Y}} = \underset{\text{fixed charge}}{1.50} + \underset{\text{hourly charge}}{2.0}\,\overset{\text{the number of hours rented}}{X}.$$

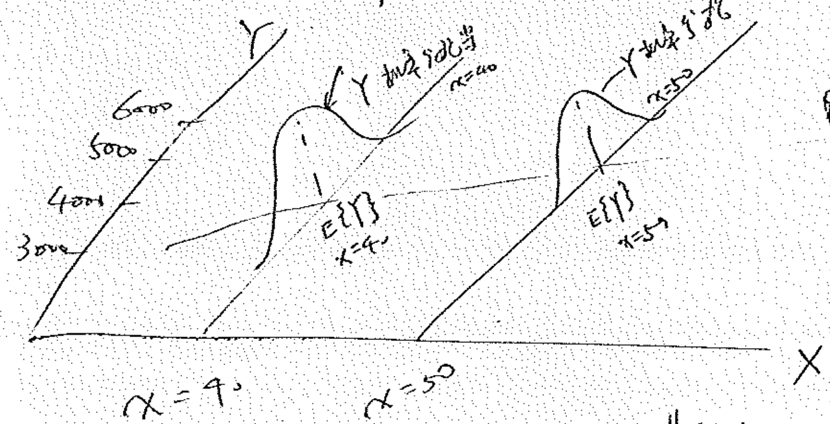
A statistical relation between two variables X and Y is not exact,

the value of Y is not uniquely determined when the value is specified.



In most empirical studies, the value of Y is not uniquely determined when the level of the independent variable is specified.

$$\overset{R.V.}{Y_i} = \underbrace{\beta_0 + \beta_1 \overset{R.V}{X_i}}_{\text{line of statistical relationship component}} + \underbrace{\varepsilon_i}_{\text{Random scatter component}}$$



Regression function 迴歸函数

the means of the probabilities distributions are located on the regression function.

假设 ① ② ③ $\varepsilon_i \sim N(\underset{①}{0}, \underset{③}{\sigma^2})$ ← random scatter balance around the line

$P(E_1 \cap E_2) = P(E_1) \cdot P(E_2)$

② ε_i for different observations are statistically independent.

$$\overset{R.V.}{Y_i} = \beta_0 + \beta_1 \overset{R.V.}{X_i} + \varepsilon_i \quad, i = 1, 2 \ldots n \quad \leftarrow \text{simple linear regression}$$

regression model

Y_i = 第 i case 的 response.

X_i = the value of the independent variable in the ith case, assume to

$E\{Y_i\} = E\{\beta_0 + \beta_1 X_i + \varepsilon_i\} = \beta_0 + \beta_1 X_i + E\{\varepsilon_i\}$ since $E\{\varepsilon_i\} = 0$

$E\{Y_i\} = \beta_0 + \beta_1 X_i \Rightarrow \boxed{\text{regression function}}$ 迴歸函數 ☆

平均值 of Y.

∴ 就 model $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ 而言，其迴歸函數

$E\{Y\} = \beta_0 + \beta_1 X$ 就任何 X 值而言，X 與 $E\{Y\}$ 之間

的關係稱為迴歸函數 β_0、β_1 為迴歸參數，β_0 截距，β_1 斜率.

由前面已知 $E\{Y_i\} = \beta_0 + \beta_1 X_i$ 再來要找出 Variance of Y_i

即然 $\beta_0 + \beta_1 X_i$ 是常數 $\sigma^2\{Y_i\} = \sigma^2\{\beta_0 + \beta_1 X_i + \varepsilon_i\} = \sigma^2\{\varepsilon_i\}$

$\sigma^2\{\varepsilon_i\} = \sigma^2$，於是知道 $\sigma^2\{Y_i\} = \sigma^2$ 誤差項之變異數

於是 Y_i 有相同之變異，不論 X 值為何，與誤差項有關

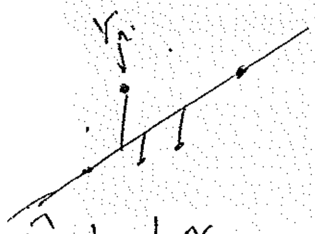
1. $Y_i \sim$ Normal 2. $E\{Y_i\} = \beta_0 + \beta_1 X_i$ 3. $\sigma^2\{Y_i\} = \sigma^2$

$\Rightarrow$ Y_i are independent $N(\overset{\text{mean}}{\beta_0 + \beta_1 X_i}, \overset{\text{variance}}{\sigma^2})$ $i = 1, 2, \ldots, n$

有人用這種方式表示

一般而言，β_0 和 β_1 是未知，必需由樣本資料估計。

最常用的方法是最小平方法 the method of least square.



找出一條直線使其(deviation of Y 觀察值和直線)^2 之平方和最小.

$\hat{Y} = b_0 + b_1 X$

∵ b_0, b_1 代表 β_0 β_1 之估計值。

令 Q 代表 the sum of the squared deviations.
離差

$$Q=\sum_{i=1}^{n}\left[Y_i-(b_0+b_1X_i)\right]^2$$

$$\left.\begin{aligned}\frac{\partial Q}{\partial b_0}&=\frac{\partial\sum_{i=1}^{n}\left[Y_i-(b_0+b_1X_i)\right]^2}{\partial b_0}=-2\sum_{i=1}^{n}(Y_i-b_0-b_1X_i)=0\\ \frac{\partial Q}{\partial b_1}&=\frac{\partial\sum_{i=1}^{n}\left[Y_i-(b_0+b_1X_i)\right]^2}{\partial b_1}=-2\sum_{i=1}^{n}X_i(Y_i-b_0-b_1X_i)=0\end{aligned}\right\}$$

$$\Rightarrow\begin{cases}\sum_{i=1}^{n}(Y_i-b_0-b_1X_i)=0\\ \sum_{i=1}^{n}X_i(Y_i-b_0-b_1X_i)=0\end{cases}\Rightarrow\begin{aligned}&\sum_{i=1}^{n}Y_i-nb_0-b_1\sum_{i=1}^{n}X_i=0\\ &\sum_{i=1}^{n}X_iY_i-b_0\sum_{i=1}^{n}X_i-b\sum_{i=1}^{n}X_i^2=0\end{aligned}$$

$$\Rightarrow\left.\begin{aligned}&\sum_{i=1}^{n}Y_i=nb_0+b_1\sum_{i=1}^{n}X_i \quad\text{——①}\\ &\sum_{i=1}^{n}X_iY_i=b_0\sum_{i=1}^{n}X_i+b_1\sum_{i=1}^{n}X_i^2 \quad\text{——②}\end{aligned}\right\}$$ 有人稱 ①② 為 normal equations

解由 ①② 將可得到 b_0, b_1 得到 $\beta_0\ \beta_1$ 之估计值

由 ① 式 $b_0=\dfrac{\sum_{i=1}^{n}Y_i-b_1\sum_{i=1}^{n}X_i}{n}=\dfrac{1}{n}\left(\sum_{i=1}^{n}Y_i-b_1\sum_{i=1}^{n}X_i\right)$

或相等 $\underline{\overline{Y}-b_1\overline{X}}$
另一個

將 b_0 代入 ② 式 可得

$$\sum_{i=1}^{n}X_iY_i=\left(\frac{\sum_{i=1}^{n}Y_i-b_1\sum_{i=1}^{n}X_i}{n}\right)\sum_{i=1}^{n}X_i+b_1\sum_{i=1}^{n}X_i^2$$

$$\sum_{i=1}^{n}X_iY_i=\frac{\sum_{i=1}^{n}Y_i\sum_{i=1}^{n}X_i-b\left(\sum_{i=1}^{n}X_i\right)^2}{n}+b_1\sum_{i=1}^{n}X_i^2$$

$$\sum_{i=1}^{n}X_iY_i=\frac{\sum_{i=1}^{n}Y_i\sum_{i=1}^{n}X_i}{n}-b_1\left[\frac{\left(\sum_{i=1}^{n}X_i\right)^2}{n}-\sum_{i=1}^{n}X_i^2\right]$$

$$\therefore\ b_1 = \frac{\frac{\sum_{i=1}^{n} Y_i \sum_{i=1}^{n} X_i}{n} - \sum_{i=1}^{n} X_i Y_i}{\frac{(\sum_{i=1}^{n} X_i)^2}{n} - \sum_{i=1}^{n} X_i^2} \Rightarrow b_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

乘一負號 ⇓　　　　　　　　　　　　　　　　　　　另一個

$$= \frac{\sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} Y_i \sum_{i=1}^{n} X_i}{n}}{\sum_{i=1}^{n} X_i^2 - \frac{(\sum_{i=1}^{n} X_i)^2}{n}}$$

上下同乘以 n ⇒ 則得到講義中之公式.

$E\{Y\} = \beta_0 + \beta_1 X$ 迴歸函數是以

$\hat{Y} = b_0 + b_1 X$ 估計 → 稱之為估計迴歸函數

$\hat{Y}_h = b_0 + b_1 X_h$

當 $X_h = 4$ 代入可得 $\hat{Y}_h$ ⇒ 是為 $E\{\hat{Y}_h\} = \beta_0 + \beta_1 X_h = E\{Y_h\}$

Series Time

平均那段時間　　不偏估計值

residual $= Y_i - \hat{Y}_i = e_i$ 以 e_i denote.

ε_i 通常是未知, 由 e_i 去估計.

the deviation of Y_i from the true mean $E\{Y_i\}$

參考書目:

John Neter, William Wasserman, Michael H. Kutner, (1990), Applied linear statistical models: regression, analysis of variance, and experimental designs, Homewood, IL : Irwin, c1990.

Bowerman, B. L. and O'Connel, R. T. (1993), *Forecasting and Time Series: An Applied Approach*, 3rd Edition, CA: Duxbury Press.Box, G. E. P. and Jenkins, G. M. (1976), *Time Series Analysis, Forecasting and Control*, 2nd Edition, San Francisco: Holden-Day.