第八屆校際系統建模與優化競賽

學生講座系列二

圖論的應用(Graph Applications) 、 Solver 應用及線性迴歸 (Linear Regression)

4th May, 2013

圖論的應用

1) 最小生成樹問題 (Minimum Spanning Tree Problem)

(I) **如果有 n 個頂點，我們（只能）用(n-1)條邊，使任意兩點相通，就是一棵生成樹。**

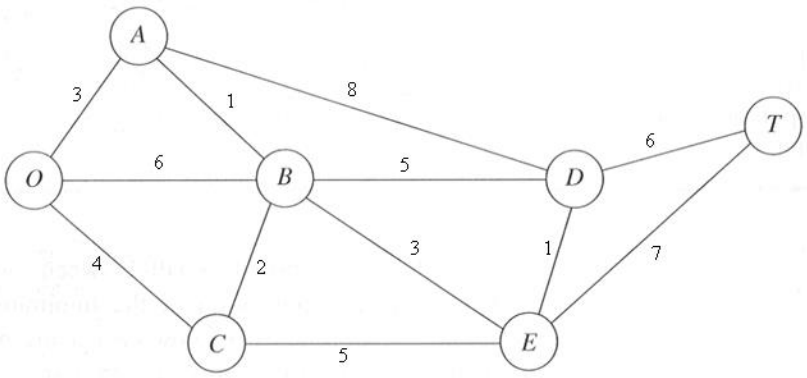
(II) **假設每一點與其他點相連的邊給予一長度，你用最少的總長度把 n 點組成一棵生成樹，就是最少生成樹。**

(III) **最小生成樹可以不唯一。**

(IV) **用途主要是在於電訊網絡（如舖設光纖、電話線）或舖路等。**

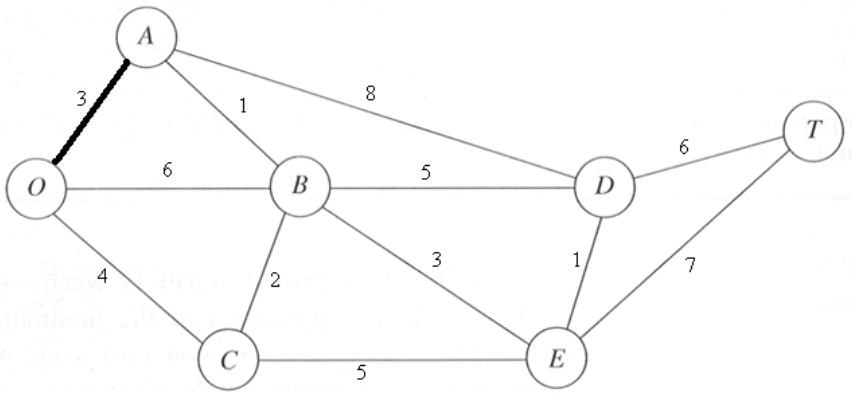
(ST1) 例題：

電話公司正考慮為一個新市鎮舖設電話線。鎮內有七個屋苑，每兩個屋苑間的舖線費用如下（單位以十萬計）：

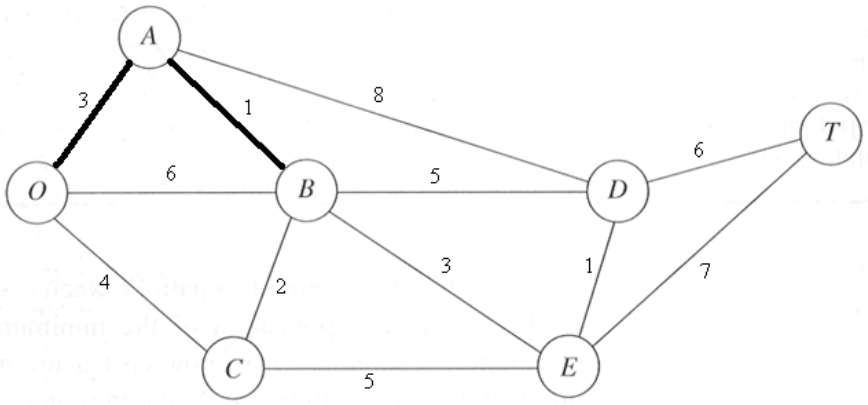


請以最少成本為該市鎮舖設電話線，使鎮內所有屋苑能互相連絡。

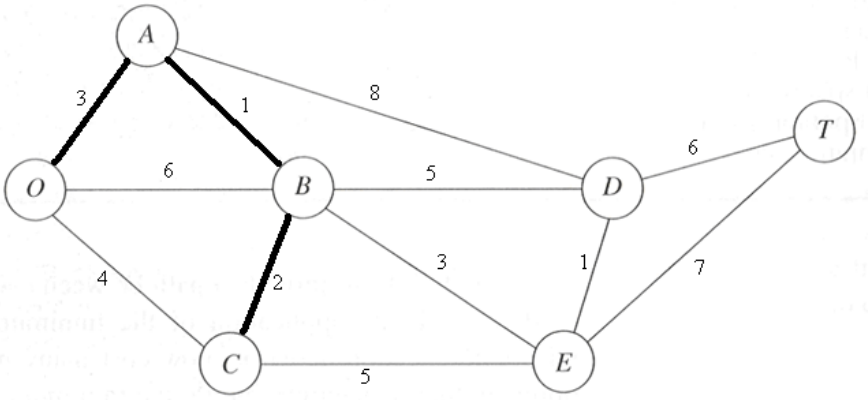
解： (i) 隨意以 O 為起點，在相鄰屋苑選取最便宜的舖設費，即 A。



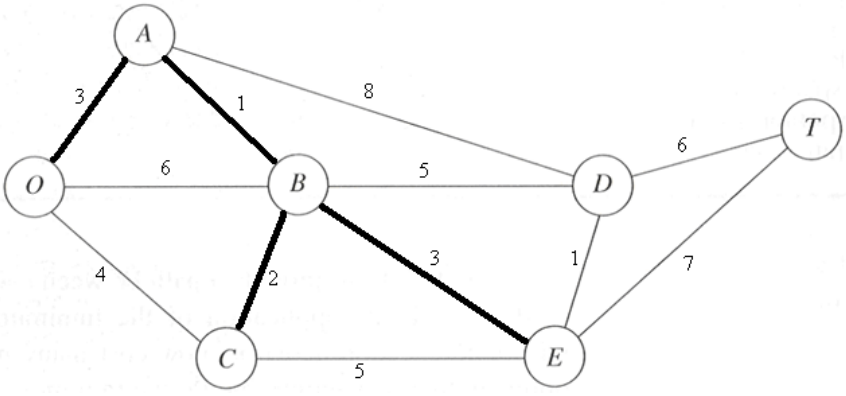
(ii) 現在 O 和 A 已連繫。在它們相鄰的屋苑再選取最便宜的舖設費，即選由 A 到 B。



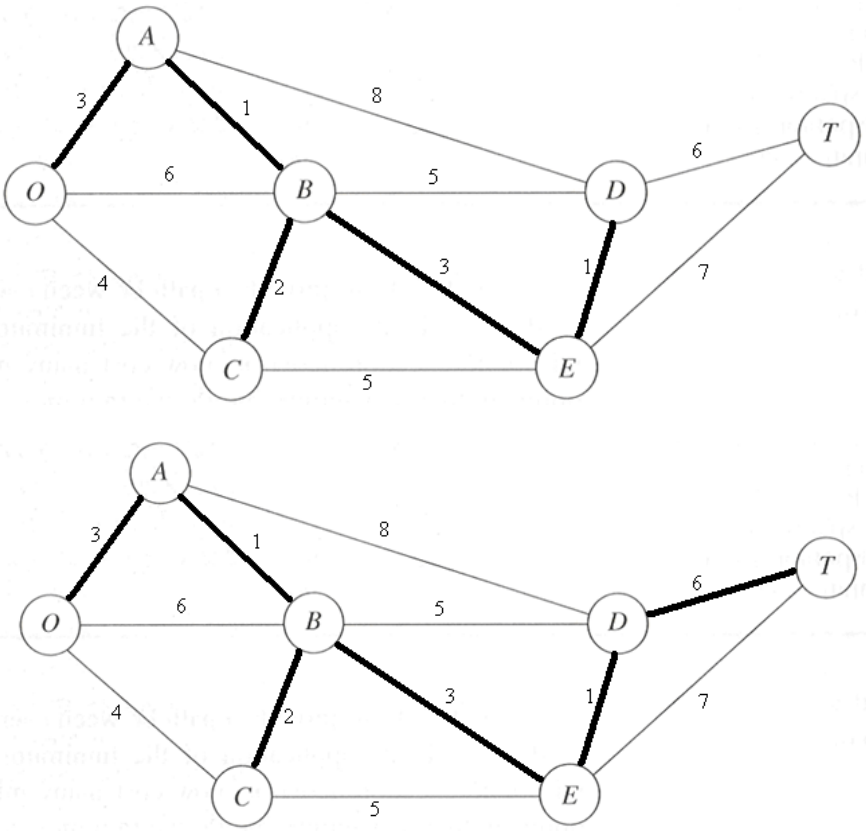
(iii) 現在 O、A 和 B 已連繫。在它們相鄰的屋苑再選取最便宜的舖設費，即選由 B 到 C。



(iv) 現在 O、A、B 和 C 已連繫。在它們相鄰的屋苑再選取最便宜的舖設費，即選由 B 到 E。

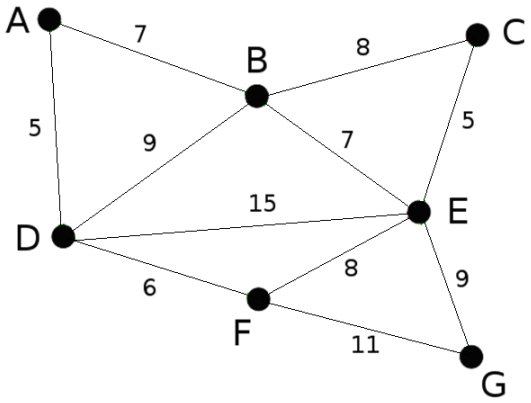


(v)　如此類推，我們會先選由 E 到 D，再選由 D 到 T。

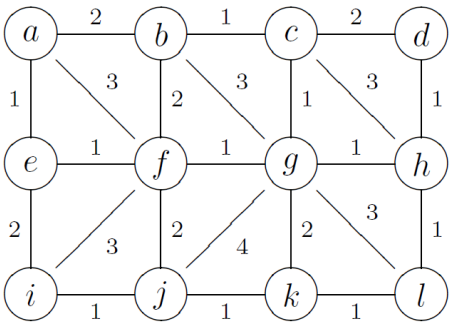


(vi)　這時所有屋苑已經互相連繫，且達到最小舖設成本，即合共一百六十萬元。(3+1+2+3+1+6)

(ST2) 系工寬頻計劃為中大花園鋪設光纖，下圖為各座之間的鋪設費用。試為系工草擬鋪設藍圖，使其成本減至最低。



(ST3) 一所電車公司計劃拓展新路線，預計為有 12 個站，下圖列出各站之間鋪設路軌的費用。試為電車公司訂出路線圖，使鋪設路軌成本降至最低。（路軌必須串連所有站。）



2) 最大流量問題 (Maximum Flow Problem)

(I) **流量是對於一個網路而言。「網路」是一個加權有向圖(weighted directed graph)，而圖上有兩個特殊的點，稱為源點(source)和終點(sink)。無任何連線(directed link/edge)指向源點；亦無任何連線由終點指向其它節點。**

(II) 其他點稱為「中繼點」(transshipment nodes)

(III) 流量(flow)的方向是根據箭嘴指示。每條路線的流量不能超過該道的「容量」(capacity)。

(IV) 目標：我們希望找出由源點到終點的最大流量。

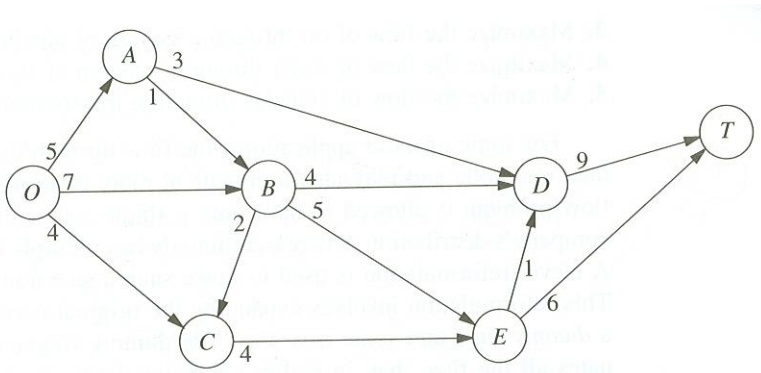
(V) 日常應用例子：

(i) 在供應網內，把最多的原材料運送到各工廠。

(ii) 在交通網絡，提高車輛流量。

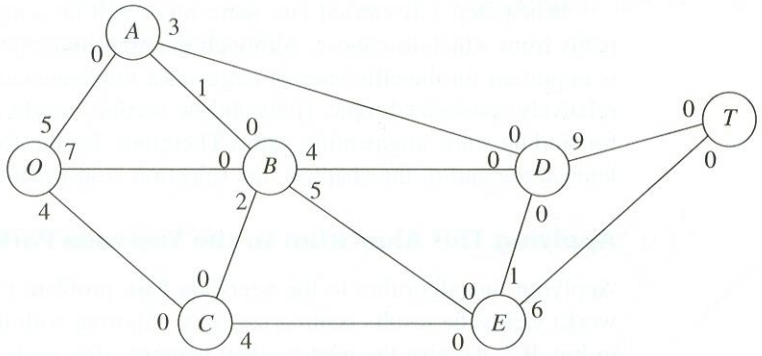
(NF1) **例題：**

在(ST1)的新市鎮內，電車公司考慮在繁忙時候安排在各站的班次，使來往 O 點和 T 點的總班次增至最多。由於電車經原路回程，我們只須考慮單方向，如下圖由 O 到 T 便可。

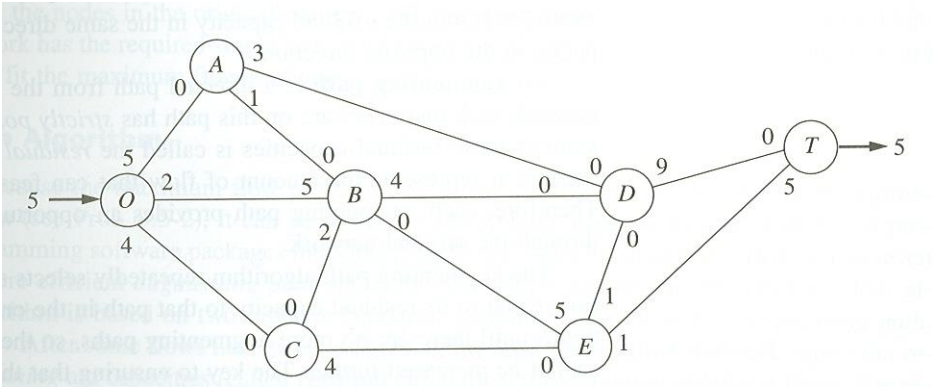


（註：OA 箭嘴上的 5，即表示由 O 往 A 最多只開出 5 班車）

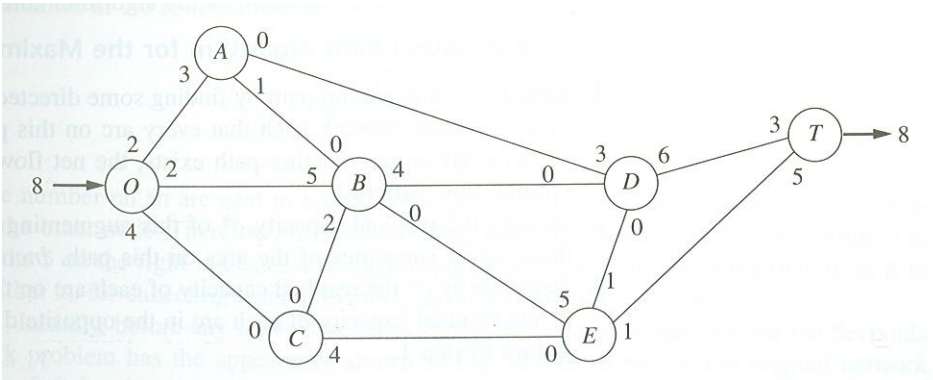
解： (i) 先把上圖化為一個沒有箭嘴的圖如下。OA 上的 5 和 0 分別表示由 O 往 A 最多有 5 班車，反方向 A 往 O 有 0 班。留意原來箭嘴的局限仍然保持。



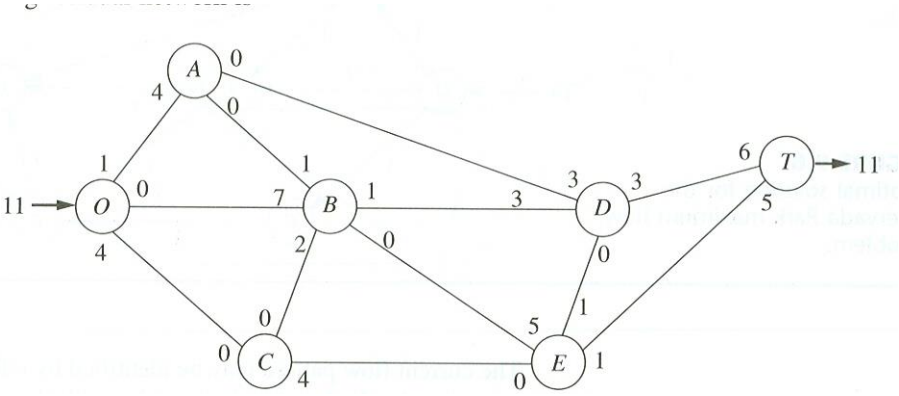
(ii) 選一條 augmenting path O→B→E→T，其最大容量（同一次）為 min{7,5,6} = 5。我們便分配 5 個單位容量到這條 augmenting path：



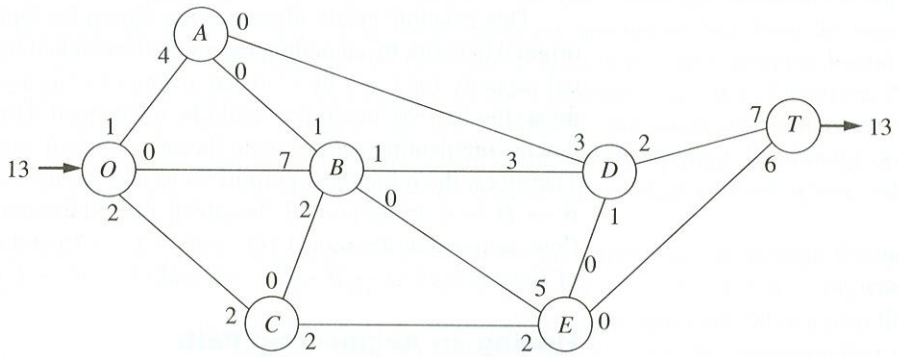
(iii) 選 augmenting path O→A→D→T，其最大容量（同一次）為 min{5,3,9} = 3。我們便分配 3 個單位容量到這條 augmenting path：



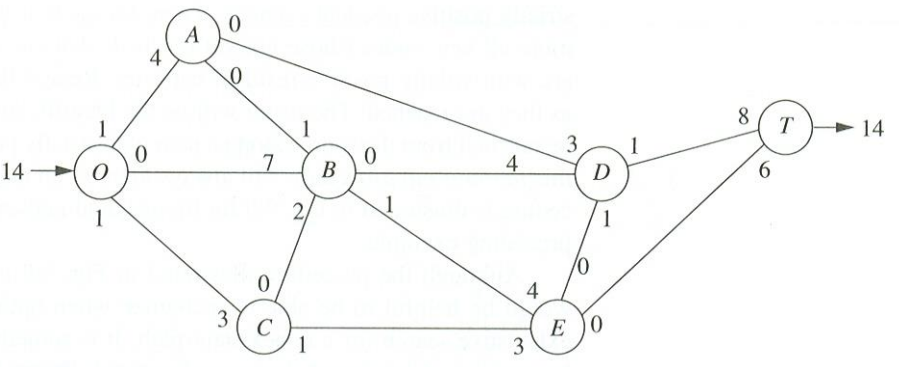
(iv) 選 augmenting path O→A→B→D→T，其最大容量（同一次）為 min{2,1,4,6} = 1。我們便分配 1 個單位容量到這條 augmenting path。再選 augmenting path O→B→D→T，其最大容量（同一次）為 min{2,4,6} = 2。我們便分配 2 個單位容量到這條 augmenting path。兩次分配後的結果：



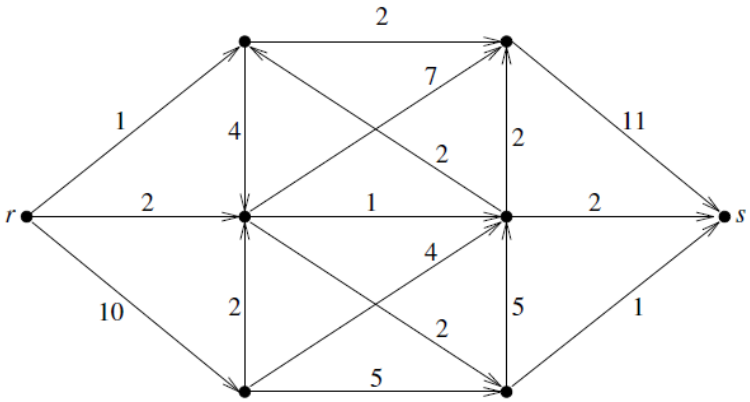
(v) 選 augmenting path O→C→E→D→T，其最大容量（同一次）為 min{4,4,1,3} = 1。我們便分配 1 個單位容量到這條 augmenting path。再選 augmenting path O→C→E→T，其最大容量（同一次）為 min{4,4,1} = 1。我們便分配 1 個單位容量到這條 augmenting path。兩次分配後的結果：



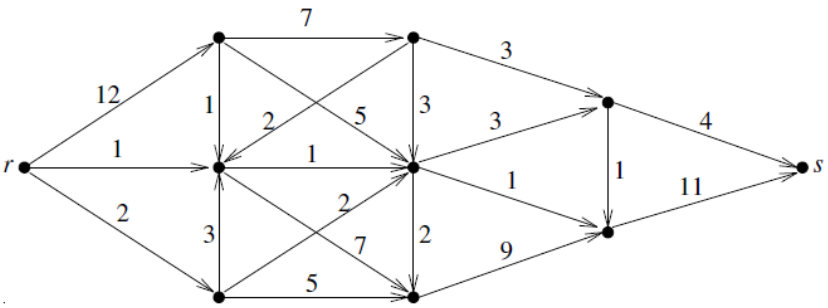
(vi) 選 augmenting path O→C→E→B→D→T，其最大容量（同一次）為 min{2,2,5,1,2,6} = 1。我們便分配 1 個單位容量到這條 augmenting path：



(NF2) 下圖為一間供水公司的地下供水網路，箭嘴代表水管，箭嘴旁的數字代表水管的最大容量。r 代表水源 ，s 代表最終用戶。試為該公 司 導 向 水流，使供水量提升至最大值。



(NF3) 在錯綜複雜的街道上，車輛應如分流，才能最有效疏導交通？



Solver 的應用

1) **建立／轉化線性規劃問題**

2) **已學課題例題**

(T1) ABC 玩具公司設有 3 間工廠，4 間門市。工廠 1、2、3 每月分別能 生產　12、17、11 萬件玩具，而每間門市每月要求入貨 10 萬件。各工廠與門市距離(公里)分別如下：

工廠 \ 門市	1	2	3	4
1	8	13	4	7
2	11	14	6	10
3	6	12	8	9

運費為每公里每件$1，求最少的運費使玩具貨品能運至各門市，並能滿足需求。

建模：

先為這十二格各設一個「分配」變數如下

工廠 \ 門市	1	2	3	4
1	8 (x_{11})	13 (x_{12})	4 (x_{13})	7 (x_{14})
2	11 (x_{21})	14 (x_{22})	6 (x_{23})	10 (x_{24})
3	6 (x_{31})	12 (x_{32})	8 (x_{33})	9 (x_{34})

每間工廠有供應限制：

$$\begin{cases} x_{11}+x_{12}+x_{13}+x_{14}=12 \\ x_{21}+x_{22}+x_{23}+x_{24}=17 \\ x_{31}+x_{22}+x_{33}+x_{34}=11 \end{cases}$$

每間門市亦有既定需求：

$$\begin{cases} x_{11}+x_{21}+x_{31}=10 \\ x_{12}+x_{22}+x_{32}=10 \\ x_{13}+x_{23}+x_{33}=10 \\ x_{14}+x_{24}+x_{34}=10 \end{cases}$$

我們的目標就是要降低運費。如果工廠一運 x_{13} 件貨到門市三，則運費須要 $4x_{13}$ 元。所以總運費為

$$\begin{aligned} &8x_{11}+13x_{12}+4x_{13}+7x_{14} \\ &+11x_{21}+14x_{22}+6x_{23}+10x_{24} \\ &+6x_{31}+12x_{32}+8x_{33}+9x_{34} \end{aligned}$$

所以我們得出以下模型：

$$
\begin{aligned}
\min \quad & 8x_{11}+13x_{12}+4x_{13}+7x_{14} \\
& +11x_{21}+14x_{22}+6x_{23}+10x_{24} \\
& +6x_{31}+12x_{32}+8x_{33}+9x_{34} \\
s.t. \quad & \begin{cases}
x_{11}+x_{12}+x_{13}+x_{14}=12 \\
x_{21}+x_{22}+x_{23}+x_{24}=17 \\
x_{31}+x_{22}+x_{33}+x_{34}=11 \\
x_{11}+x_{21}+x_{31}=10 \\
x_{12}+x_{22}+x_{32}=10 \\
x_{13}+x_{23}+x_{33}=10 \\
x_{14}+x_{24}+x_{34}=10 \\
x_{11},x_{12},x_{13},x_{14},x_{21},x_{22},x_{23},x_{24},x_{31},x_{32},x_{33},x_{34}\geq 0
\end{cases}
\end{aligned}
$$

(A1) 列印工作安排

小明要打印大量不同大小的文件，其中包括有 A4、Letter 及 Legal。他有三台不同性能的印表機，每分鐘的印紙量如下：

	A4	Letter	Legal
甲機	10 張	12 張	11 張
乙機	8 張	10 張	9 張
丙機	12 張	14 張	12 張

假設小明有無限的紙和無限的文件要打印，但小明每款紙匣只有一個，每台印表機只可同時打印一種紙。

1) 請問小明該如何安排打印工作，以求用最少的時間，打印最多的文件？

建模：

先為這九格各設一個「分配」變數如下

	A4	Letter	Legal
甲機	10 張(x_{11})	12 張(x_{12})	11 張(x_{13})
乙機	8 張(x_{21})	10 張(x_{22})	9 張(x_{23})
丙機	12 張(x_{31})	14 張(x_{32})	12 張(x_{33})

如甲機被分配去印 A4 紙，則 $x_{11}=1$ ，$x_{12}=0$ ，$x_{13}=0$。各機如是，所以：

而每款紙匣只有一個，例如 Letter 紙匣被乙機去了，則 $x_{21}=0$ ，$x_{22}=1$ ，$x_{23}=0$。各紙匣如是，所以：

$$\begin{cases} x_{11}+x_{21}+x_{31}=1 \\ x_{21}+x_{22}+x_{32}=1 \\ x_{31}+x_{32}+x_{33}=1 \end{cases}$$

如果把「分配」變數乘以對應格內的張數，可理解如下：

看$10x_{11}$，若甲機真的印 A4 紙，乘積是 10，否則是 0。

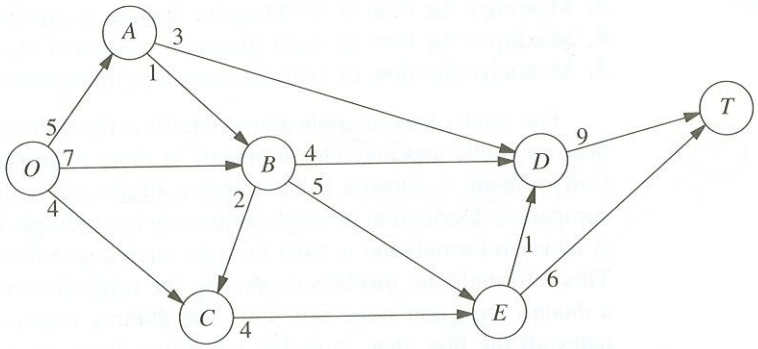
小明希望列印總數最多，可看作提升以下函數至最大值，也就是他的目標函數：

$$\begin{aligned} &10x_{11}+12x_{12}+11x_{13} \\ &+8x_{21}+10x_{22}+9x_{23} \\ &+12x_{31}+14x_{32}+12x_{33} \end{aligned}$$

所以我們得出以下模型：

$$\begin{aligned} \max \quad & 10x_{11}+12x_{12}+11x_{13} \\ & +8x_{21}+10x_{22}+9x_{23} \\ & +12x_{31}+14x_{32}+12x_{33} \\ s.t. \quad & \begin{cases} x_{11}+x_{12}+x_{13}=1 \\ x_{21}+x_{22}+x_{23}=1 \\ x_{31}+x_{32}+x_{33}=1 \\ x_{11}+x_{21}+x_{31}=1 \\ x_{12}+x_{22}+x_{32}=1 \\ x_{13}+x_{23}+x_{33}=1 \\ x_{11},x_{12},x_{13},x_{21},x_{22},x_{23},x_{31},x_{32},x_{33}\geq 0 \end{cases} \end{aligned}$$

(NF1) 在(ST1)的新市鎮內，電車公司考慮在繁忙時候安排在各站的班次，使來往 O 點和 T 點的總班次增至最多。由於電車經原路回程，我們只須考慮單方向，如下圖由 O 到 T 便可。



建模：

設變數 F 為總流入量（或流出量），及 f_{ij} 為 i 點到 j 點的流量。如 O 點到 A 點的流量為 1，則 $f_{OA}=1$。

根據流量的恆守(conservation of flow)，每一點的流入量和流出量應該一樣。圖中有 7 個點，因此我們有以下 7 個限制：

$$\begin{cases} f_{OA}+f_{OB}+f_{OC}=F \\ f_{OA}-f_{AD}-f_{AB}=0 \\ f_{OB}+f_{AB}-f_{BC}-f_{BD}-f_{BE}=0 \\ f_{Oc}+f_{BC}-f_{CE}=0 \\ f_{AD}+f_{BD}+f_{ED}-f_{DT}=0 \\ f_{BE}+f_{CE}-f_{ED}-f_{ET}=0 \\ f_{DT}+f_{ET}=F \end{cases}$$

而每一管道有容量上限，且不許倒流，因此每個 f_{ij} 都有上限和下限。例如對應 OA、OB、OC，我們有以下條件：

$$\begin{cases} 0 \le f_{OA} \le 5 \\ 0 \le f_{OB} \le 7 \\ 0 \le f_{OC} \le 4 \end{cases}$$

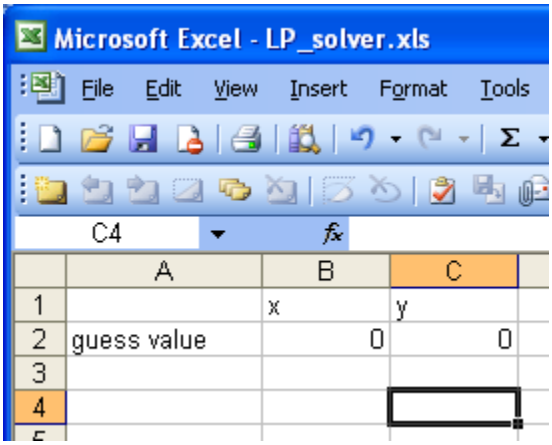
所以我們得出以下模型：

$$\max \quad F$$

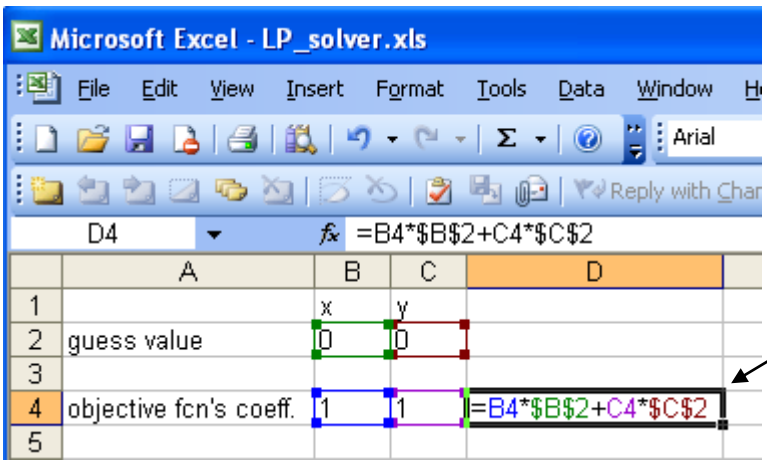
$$s.t. \quad \begin{cases} f_{OA} + f_{OB} + f_{OC} = F \\ f_{OA} - f_{AD} - f_{AB} = 0 \\ f_{OB} + f_{AB} - f_{BC} - f_{BD} - f_{BE} = 0 \\ f_{Oc} + f_{BC} - f_{CE} = 0 \\ f_{AD} + f_{BD} + f_{ED} - f_{DT} = 0 \\ f_{BE} + f_{CE} - f_{ED} - f_{ET} = 0 \\ f_{DT} + f_{ET} = F \\ 0 \le f_{OA} \le 5 \\ 0 \le f_{OB} \le 7 \\ 0 \le f_{OC} \le 4 \\ 0 \le f_{AB} \le 1 \\ 0 \le f_{AD} \le 3 \\ 0 \le f_{BC} \le 2 \\ 0 \le f_{BD} \le 4 \\ 0 \le f_{BE} \le 5 \\ 0 \le f_{CE} \le 4 \\ 0 \le f_{DT} \le 9 \\ 0 \le f_{ED} \le 1 \\ 0 \le f_{ET} \le 6 \end{cases}$$

3) **使用 Solver 一般步驟：**

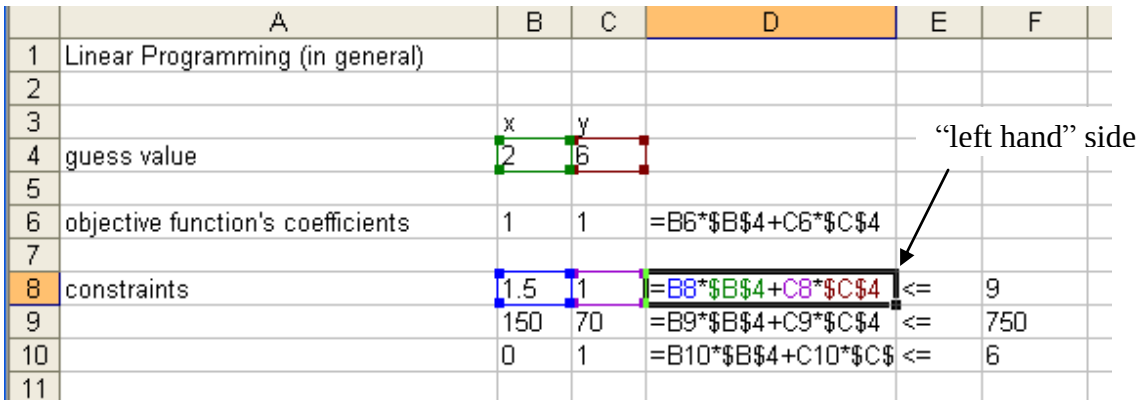
(I) 確立變數格



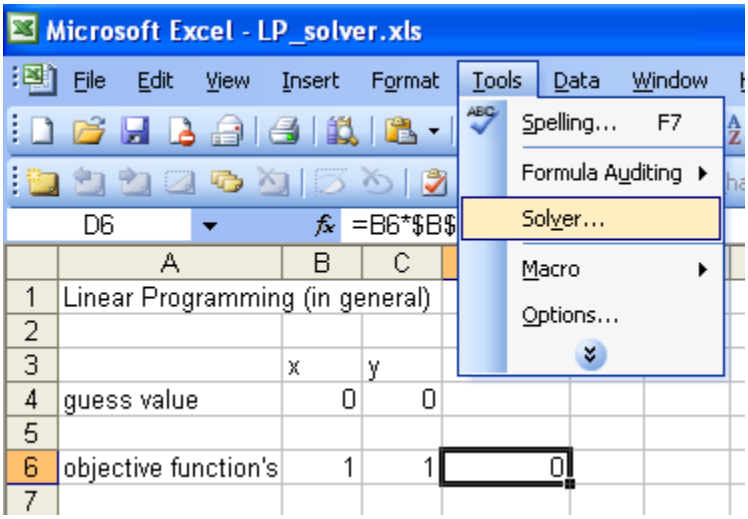
(II) 寫下目標函數的系數 (coefficients)，並訂下目標函數格。



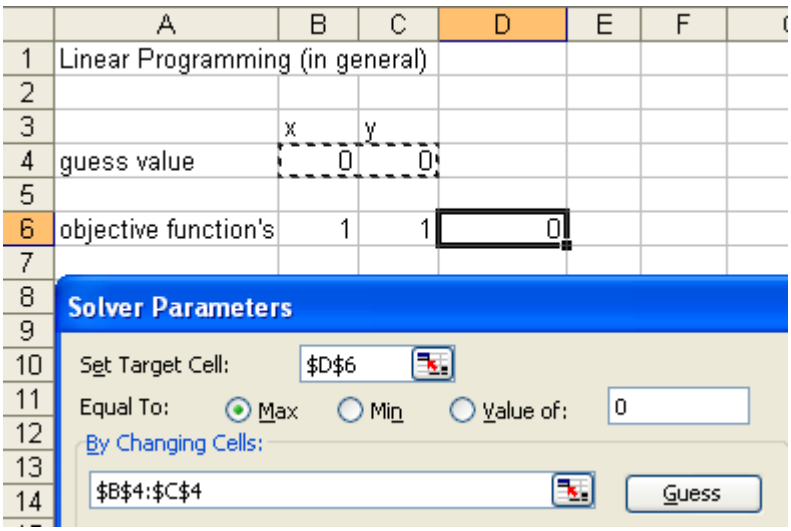
(III) 寫下各限制式的系數，並訂下各限制式的「左邊」。（暫時忽略非負數的條件。）



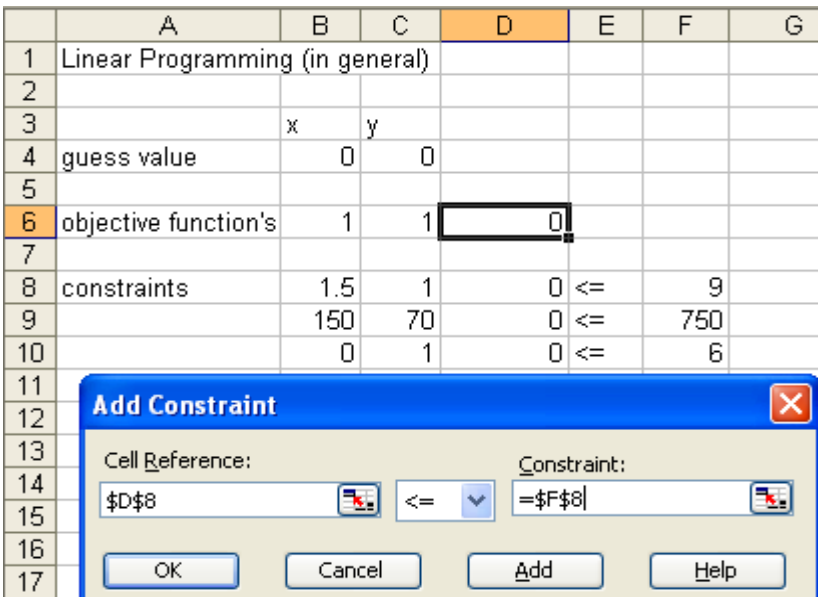
(IV) 點選目標函數格，然後在「工具」選取 “Solver”



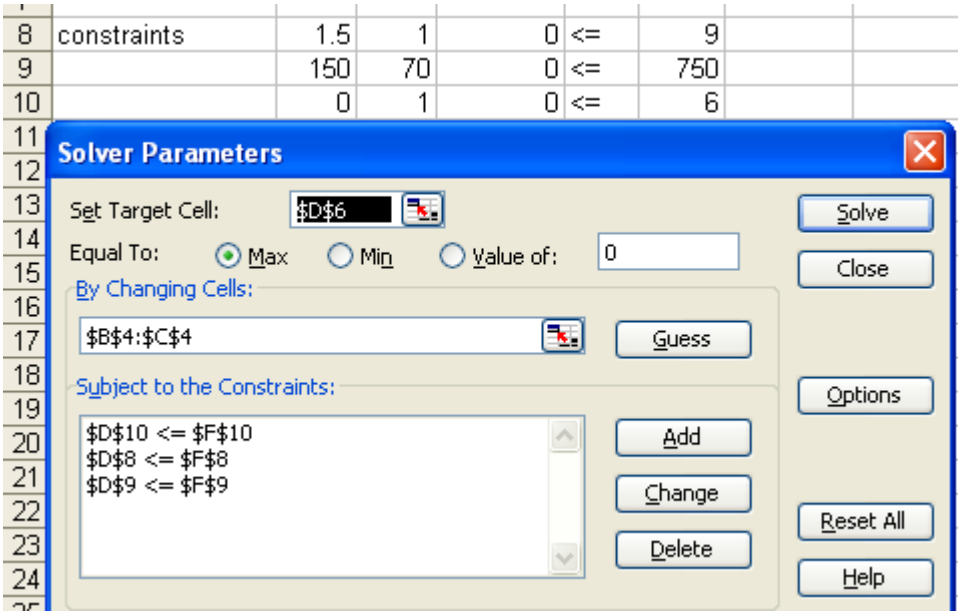
(V) 按需要選擇 “max”或”min”，在 “By changing cells”選定變數格



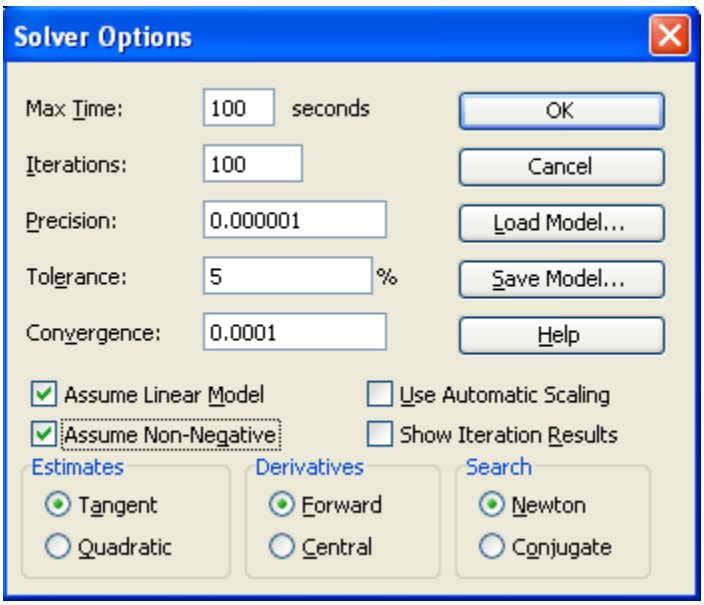
(VI) 在 Solver 上按 “Add”加入各限制式。在 “Call Reference”選取剛才定下的 “左邊”式， “Constraint”格就是對應的 “右邊”。按 OK。



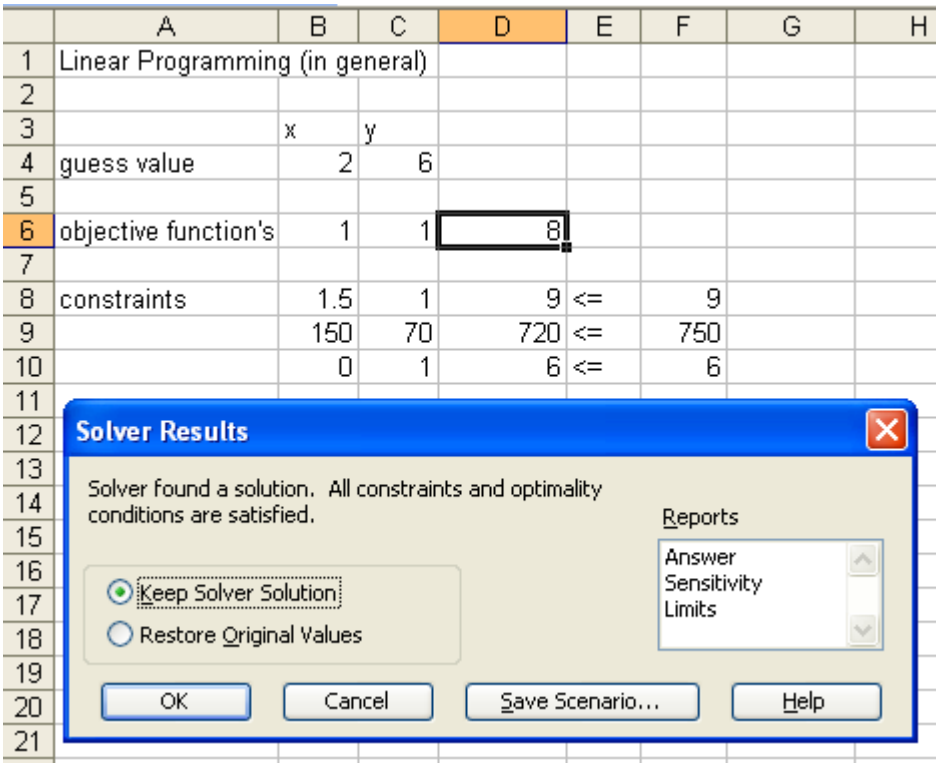
(VII) 如是者加入所有限制式：



(VIII) 在 “Option”內再加入非負數的條件，按 OK。



(IX) 再按 “Solve”，便能得出答案。可按 “Keep Solver Solution” 保留答案。



4) 之前能寫成線性規劃模型的問題，皆能用以上步驟和 Solver 解決。然而，我們學過的分配問題、交通問題和最大流量問題都有比較方便的設立方法來使用 Solver。

	A	B	C	D	E	F	
1	**Transportation Problem**						
2							
3	Variable						
4	0	0	0	0			
5	0	0	0	0			
6	0	0	0	0			
7							
8	Target coefficient						
9	8	13	4	7			
10	11	14	6	10			
11	6	12	8	9			
12							
13	Product in target						
14	=A9*A4	=B9*B4	=C9*C4	=D9*D4			
15	=A10*A5	=B10*B5	=C10*C5	=D10*D5			
16	=A11*A6	=B11*B6	=C11*C6	=D11*D6			
17			target	=SUM(A14:D16)			
18	Constraint						
19	1	1	1	1	=A19*A4+B19*B4+C19*C4+D19*D4	=	12
20	1	1	1	1	=A20*A5+B20*B5+C20*C5+D20*D5	=	17
21	1	1	1	1	=A21*A6+B21*B6+C21*C6+D21*D6	=	11
22	=A19*A4+A20*A5+A21*A6	=B19*B4+B20*B5+B21*B6	=C19*C4+C20*C5+C21*C6	=D19*D4+D20*D5+D21*D6			
23	=	=	=	=			
24	10	10	10	10			

	A	B	C	D	E	F
1	**Assignment Problem**					
2	Variable					
3	0	0	0			
4	0	0	0			
5	0	0	0	Product in target		
6				=A8*A3	=B8*B3	=C8*C3
7	Target coefficient			=A9*A4	=B9*B4	=C9*C4
8	10	12	11	=A10*A5	=B10*B5	=C10*C5
9	8	10	9		Target	=SUM(D6:F8)
10	12	14	12			
11						
12						
13	Constraints					
14	1	1	1	=A14*A3+B14*B3+C14*C3	=	1
15	1	1	1	=A15*A4+B15*B4+C15*C4	=	1
16	1	1	1	=A16*A5+B16*B5+C16*C5	=	1
17	=A14*A3+A15*A4+A16*A5	=B14*B3+B15*B4+B16*B5	=C14*C3+C15*C4+C16*C5			
18	=	=	=			
19	1	1	1			

	A	B	C	D	E	F	G	H	I	J	K
1	**Maximum Flow Problem**										
2											
3		From	To	Flow		Capacity		Nodes	Net Flow		Supply/Demand
4		O	A	3	<=	5		O	=D4+D5+D6		
5		O	B	7	<=	7		A	=-D4+D7+D8	=	0
6		O	C	4	<=	4		B	=-D5-D7+D9+D10+D11	=	0
7		A	B	0	<=	1		C	=-D6-D9+D12	=	0
8		A	D	3	<=	3		D	=-D8-D10+D13-D14	=	0
9		B	C	0	<=	2		E	=-D11-D12+D14+D15	=	0
10		B	D	4	<=	4		T	=-D13-D15		
11		B	E	3	<=	5					
12		C	E	4	<=	4					
13		D	T	8	<=	9					
14		E	D	1	<=	1					
15		E	T	6	<=	6					
16											
17	Maximum Flow	=		=I4							
18											

線性迴歸

1) 基本統計知識

(i) **mean** $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

(ii) **sample variance** $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

population variance $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$

(iii) **sample s.d.** $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$

population s.d. $\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$

2) 統計學上，我們常常探討兩樣（或多樣）東西之間的關係。我們可抽取樣本及量化為數據，看看兩組數據有沒有關係。如果有關係，我們往往會追問它們有否線性關係，因此便會使用線性迴歸分析。

3) 所謂「迴歸」，就是探討一個變數對另一變數的影響。

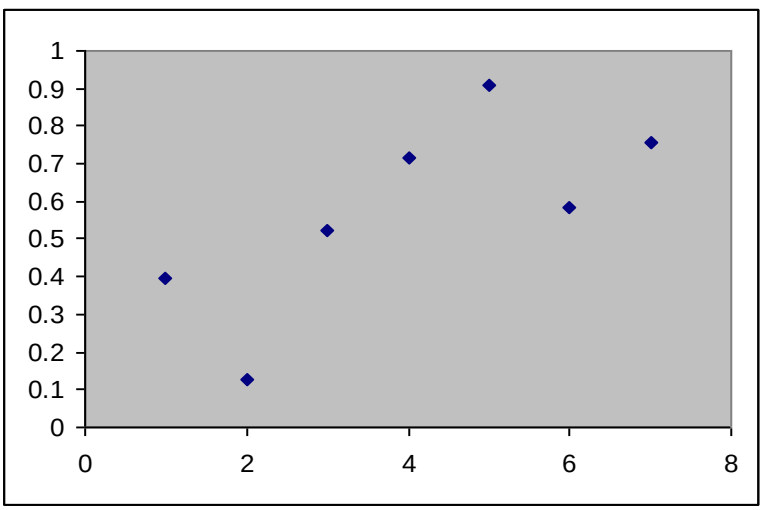
4) 模型構作

(I) 假設我們想研究 **x** 和 **y** 之間的線性關係，即想找參數**(parameter)** **α** 和 **β**使以下關係成立：

$$y = \alpha + \beta x$$

(II) 為找 **α** 和 **β**，我們需要收集一些 **x** 和 **y** 的抽本。假設我們找到了 **n** 對樣本：$(x_1, y_1), (x_2, y_2), \cdots (x_n, y_n)$

(III) 如果把這些樣本用 **xy-plane** 表示，我們會見到一些離散的點在圖上：



(IV) 找 α 和 β就相對於在圖中找一條「最好」的直線，使每點與該直線的距離最短。數學上，我們可用「最小二乘法」。

(V) 最小二乘法 **(Least Squares Method)**

如果把每對樣本(x_i, y_i)放入直線方程，應該會有一定的誤差ε_i：

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

如果把這些誤差的平方加起來，我們得出

$$L = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2$$

所謂 **"Least Square"**，就是找出找 α 和 β，使 **L**「最小」。

由於解釋過程涉及微積分，就此略過，並只列出結果。

首先定義兩個項：

$$SS_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}$$

$$SS_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}$$

當中 $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ 及 $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ 。則 α 和 β 的估值如下：

$$\begin{cases} \hat{\beta} = \dfrac{SS_{xy}}{SS_{xx}} \\ \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \end{cases}$$

(VI) 其實找到一條直線後，我們尚要透過一些統計測試，如假設檢定 **(hypothesis testing)** ，才能驗證是否合理。

(VII) 如果我們可以想了解這個線性模型究竟是否適合，我們可看看「判定係數值」**(coefficient of determination)** 或 $\mathbf{R^2}$：

首先定義兩個項：

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

及 $SS_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2$

則 $R^2 = \dfrac{SS_{yy} - SSE}{SS_{yy}} = 1 - \dfrac{SSE}{SS_{yy}}$

R^2 的意思就是看看利用這個線性模型，x 能解釋 y 多少。R^2 為 0 至 1 之間的數字，如果 R^2=0.95，表示 x 能解釋 95%的 y，所以這樣線性關係頗合理。

5) **注意：即使我們能為 x 和 y 找到一線性關係，亦不表示我們找到一個因果關係 (causal relation) 。例如，我們相信鞋帶越長，智商(IQ)便越高。這是因為當人越長大／高，所需穿的鞋的尺碼亦增大，因而需要一對較長的鞋帶。但是，你相信自己馬上去更換一對較長的鞋帶後，能使你變得更聰明嗎？**

(LR1) 一間家品店想研究宣傳 (advertising) 商品是否有助增加利潤 (sales revenue)。它抽取了最近五個月的數據 ：

Month	Advertising Expenditure x (in hundreds of dollors)	Salles Revenue y (in thousands of dollors)
1	1	1
2	2	1
3	3	4
4	4	4
5	5	8

(a) 為兩組數據找出最優擬合線 (best-fit line)。

(b) 找出該線的判定係數值。

(c) 利用(a)估算當這間家品店花 6 百元作為宣傳費時的利潤。

解(a)：$\sum_{i=1}^{5} x_i = 15$ ，$\sum_{i=1}^{5} y_i = 18$ ，$\sum_{i=1}^{5} x_i^2 = 55$ ，$\sum_{i=1}^{5} x_i y_i = 71$

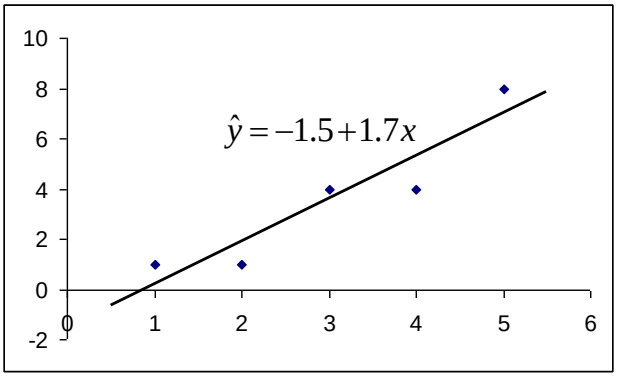
$$SS_{xy} = \sum_{i=1}^{5} x_i y_i - \frac{\left(\sum_{i=1}^{5} x_i\right)\left(\sum_{i=1}^{5} y_i\right)}{5} = 71 - \frac{(15)(18)}{5} = 17$$

$$SS_{xx} = \sum_{i=1}^{5} x_i^2 - \frac{\left(\sum_{i=1}^{5} x_i\right)^2}{5} = 55 - \frac{(15)^2}{5} = 10$$

$$\therefore \quad \hat{\beta} = \frac{SS_{xy}}{SS_{xx}} = \frac{17}{10} = 1.7$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = \frac{\sum_{i=1}^{5} y_i}{5} - \hat{\beta}\frac{\sum_{i=1}^{5} x_i}{5} = \frac{18}{5} - (1.7)\frac{15}{5} = -1.5$$

我們得出最優擬合線： $\hat{y} = -1.5 + 1.7x$



解(b)：先計算剛才定義的 SS_{yy} 和 SSE：

$$SS_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{5} y_i^2 - \frac{\left(\sum_{i=1}^{5} y_i\right)^2}{5} = 98 - \frac{(18)^2}{5} = 33.2$$

$$SSE = \sum_{i=1}^{5}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{5}(y_i - 1.5 + 1.7x_i)^2 = 4.3$$

$$\therefore \quad R^2 = 1 - \frac{1.1}{6} = 0.87$$

根據這線性模型，家品店的利潤約八成七可被宣傳費解釋。

解(c)： $\hat{y} = -1.5 + 1.7(6) = 8.7$

所以當這家品店用 6 百元宣傳費時，預計利潤可達 8700 元。

(LR2) 一間保險公司相信它的推銷員的每月推銷利潤 (monthly sales) 可隨經驗 (months on job) 而增加，以下是該公司隨意抽取的有關樣本：

Months on Job x	**Monthly Sales** y (thousands of dollars)
3	8.6
5	11.8
2	4.9
8	19.3
6	16.4
9	23.2
3	7.3
4	10.9

(a) 試為樣本找一條迴歸直線 (regression line) 。

(b) 把樣本和 (a) 部分的直線劃在圖上。

(c) 試估計一位有 9 個月經驗和一位有 6 個月經驗的推銷員的每月利潤相差多少。

(LR3) 以下記錄了 20 位同學的數學科期中試和期末試的成績：

Midterm Exam	Final Exam
85	62
52	80
84	21
43	22
85	85
71	22
88	87
81	60
87	62
76	58
54	64
91	98
77	66
71	69
75	39
95	67
86	42
82	61
40	23
80	100

(a) 請找一條最優擬合線，使我們能靠期中試的分數來估算期末試的成績。

(b) 根據這條線，估算一位期中試拿 84 分的同學的期末試分數。

(c) 計算判定係數值。根據其他科目的經驗，老師認為如果迴歸線大概能解釋七成的期末試分數才算可靠，那麼他覺得這條線可靠嗎？

6) 使用 Excel 去進行線性迴歸

當處理比較多的數據時，可以使用 Excel 內和 Solver 同樣為增益集 Data Analysis。例如金融界在分析一些股票表現的時候，住住要處理成千上萬的數據。

(LR4)附件紀錄了股票 A 和 恆生指數的 2286 日數據。 一名分析員認為股票 A 的升跌百份比應該和恆生指數的升跌百份比有線性關係。

(a) 試為分析員的意見寫出一個線性模型。
(b) 試計算出這條回歸直線。
(c) 根據這條線，如果市場(恆生指數)升幅有 4%, 估算股票 A 會升多少。

解: (a) 設 Y_t 和 X_t 分別是股票 A 在 t 日的報價和指數，

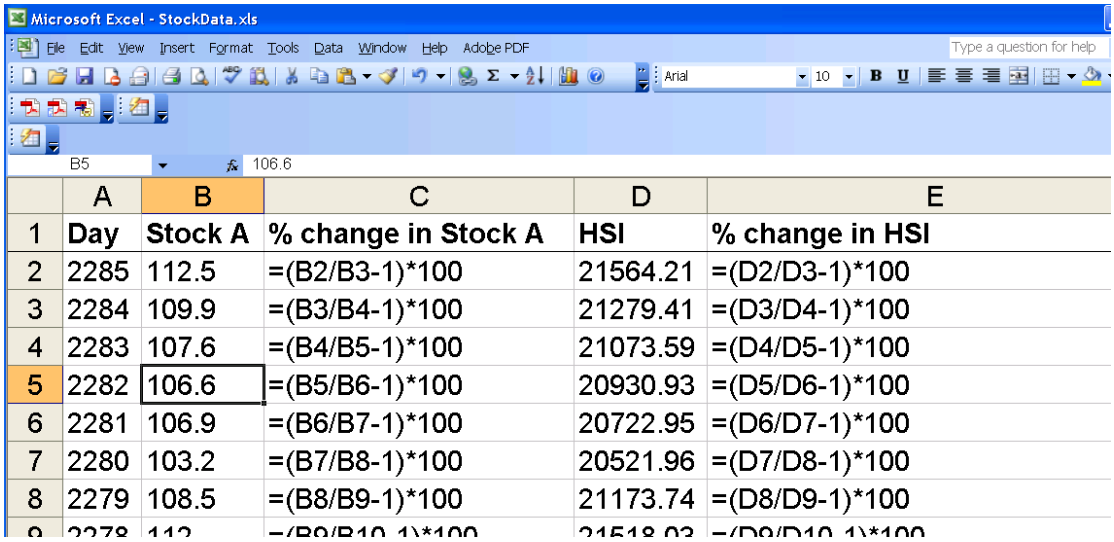
再定義$\triangle Y_t = (Y_t - Y_{t-1})/ Y_{t-1}$, $\triangle X_t = (X_t - X_{t-1})/ X_{t-1}$ 為其的升降百分比

該線性模型應是 $\triangle Y_t = a + b \triangle X_t$

(如同學想知更多關於以上金融模型，可以搜尋"CAPM Model "。)

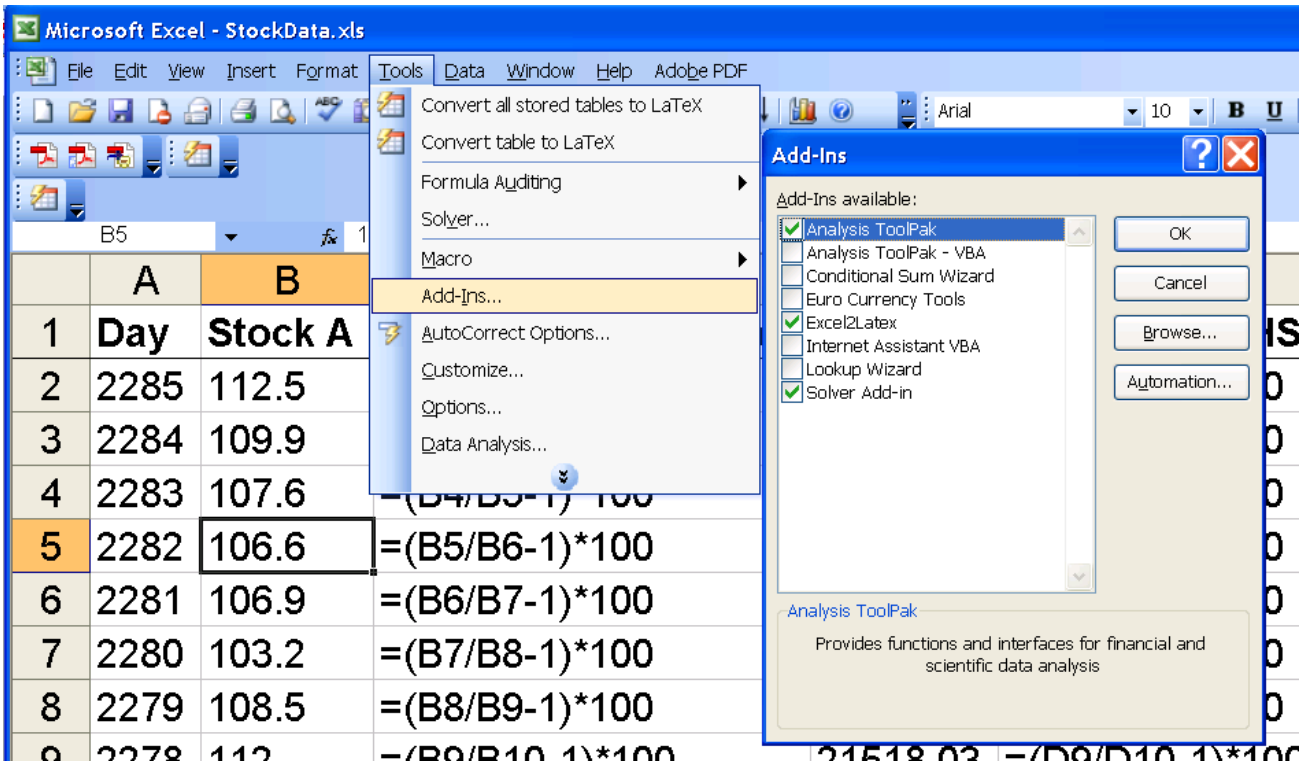
(b) 以下是利用 Data Analysis 的程序

(I) **先輸入要分析的數據，在列 C 和列 E 的算式是計算其百份比。**

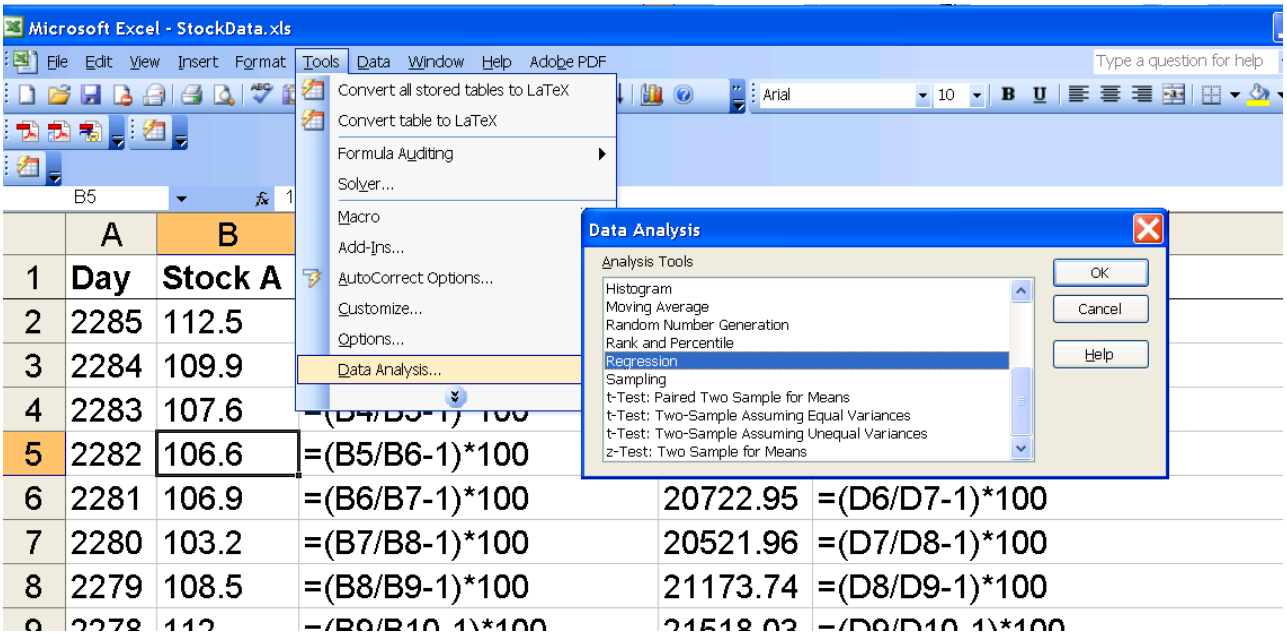


	A	B	C	D	E
1	Day	Stock A	% change in Stock A	HSI	% change in HSI
2	2285	112.5	=(B2/B3-1)*100	21564.21	=(D2/D3-1)*100
3	2284	109.9	=(B3/B4-1)*100	21279.41	=(D3/D4-1)*100
4	2283	107.6	=(B4/B5-1)*100	21073.59	=(D4/D5-1)*100
5	2282	106.6	=(B5/B6-1)*100	20930.93	=(D5/D6-1)*100
6	2281	106.9	=(B6/B7-1)*100	20722.95	=(D6/D7-1)*100
7	2280	103.2	=(B7/B8-1)*100	20521.96	=(D7/D8-1)*100
8	2279	108.5	=(B8/B9-1)*100	21173.74	=(D8/D9-1)*100
9	2278	112	=(B9/B10-1)*100	21518.03	=(D9/D10-1)*100

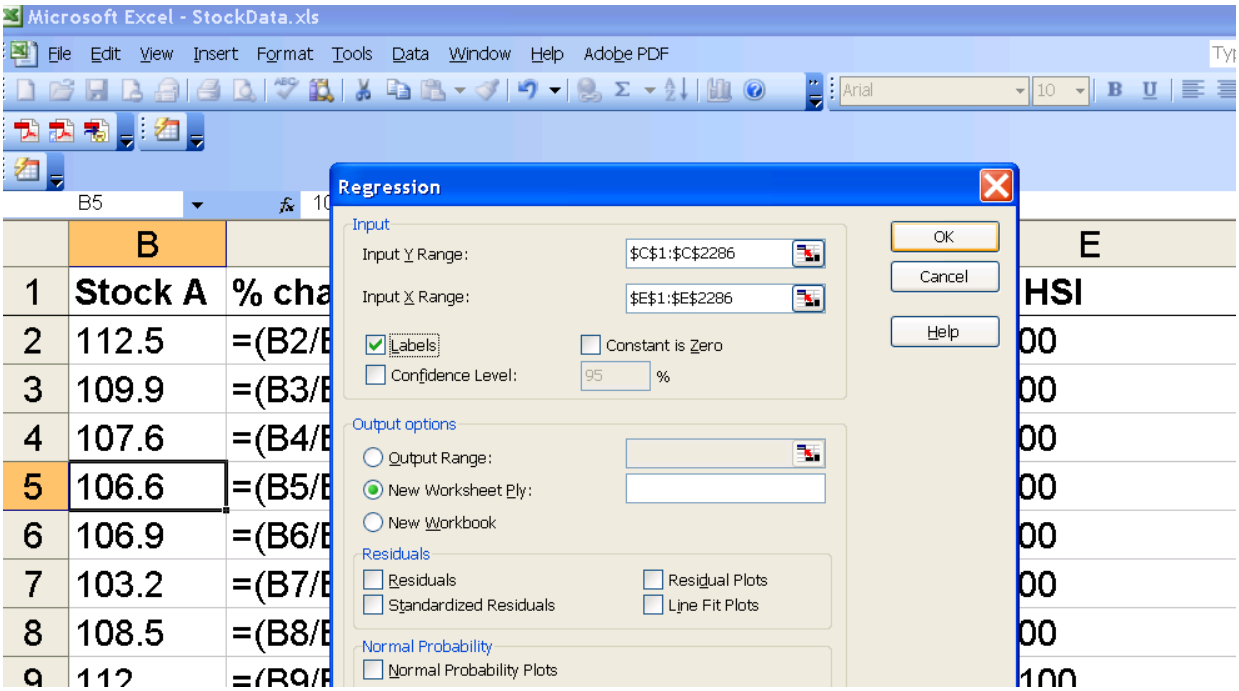
(II)　選擇 Tools → Add-Ins → 打勾 Analysis TooPak → 按 ok



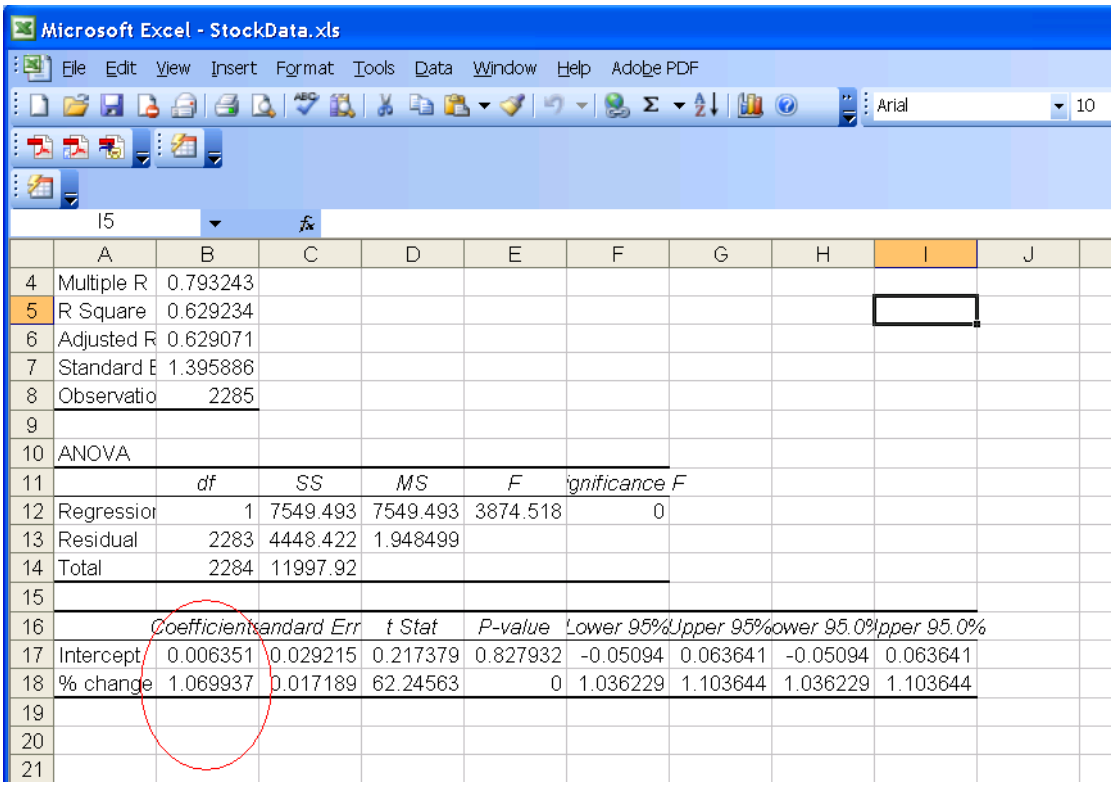
(III)　選擇 Tools → Data Analysis → Regression



(IV) 輸入$\triangle Y_t$,$\triangle X_t$ 的資料，再按 **ok**.



(V) 分析結果會自動出現， 紅色圈內的數字分別是其回歸直線的截距和斜率。



$$\Delta\hat{Y}_t = 0.00635 + 1.069937\Delta X_t$$

(c) 如果恆生指數升幅有 4%, 根據以上回歸直線，股票 A 估計會升 4.286%

YAHOO! 財經

http://hk.finance.yahoo.com

OANDA 歷史匯率

http://www.oanda.com/lang/cnt/currency/historical-rates/

香港統計資料

http://www.censtatd.gov.hk/hkstat/index_tc.jsp

資料一線通

http://www.gov.hk/en/theme/psi/datasets/

Data.gov (From US government)

http://www.data.gov/