Bài giảng

GIẢI CÁC BÀI TOÁN TỐI ƯU VÀ THỐNG KÊ TRÊN MICROSOFT EXCEL

PGS. TS. Bùi Thế Tâm
Phòng Tối ưu và Điều khiển
Viện Toán học
Viện Khoa học và Công nghệ Việt nam

Tóm tắt . Microsoft Excel 2000, 2003 có các công cụ toán học rất mạnh để giải các bài toán tối ưu và thống kê toán học. Excel có thể giải được các loại bài toán tối ưu: bài toán quy hoạch tuyến tính tổng quát, các biến có thể có ràng buộc hai phía, ràng buộc cũng có thể viết ở dạng hai phía; bài toán vận tải có hai chỉ số; bài toán quy hoạch nguyên (các biến có điều kiện nguyên hay boolean); bài toán quy hoạch phi tuyến. Số biến cúa bài toán quy hoạch tuyến tính hay nguyên có thể lên tới 200 biến. Excel còn có thể giải các bài toán hồi quy trong thống kê toán học: hồi quy đơn, hồi quy bội, hồi quy mũ.

Chương 1 có thể dạy bổ sung vào sau giáo trình *Quy hoạch tuyến tính* hay *Quy hoạch nguyên* ở bậc đại học để sinh viên có thể giải ngay trên máy tính các bài toán tối ưu cỡ lớn phát sinh trong thực tiễn mà không cần phải lập trình. *Chương 2* có thể dạy bổ sung vào sau giáo trình *Xác suất thống kê* ở bậc đại học để sinh viên có thể tính ngay được các bài toán hồi quy trên máy tính. Cả hai chương này đều có thể dạy cho sinh viên ngay sau phần Excel của môn *Tin học văn phòng*. Đây là bài giảng của tác giả cho sinh viên một số trường kinh tế và kỹ thuật.

Vài nét về tác giả. B.T.Tâm hiện làm việc tại Phòng Tối ưu và Điều khiển thuộc Viện Toán học, Viện khoa học và công nghệ Việt nam, bảo vệ Tiến sỹ năm 1978 tại Viện hàn lâm Khoa học Liên xô. Địa chỉ liên hệ: Bùi Thế Tâm, Viện Toán học, 18 Hoàng Quốc Việt, 10307 Hà Nội. Địa chỉ email: bttam@math.ac.vn. Điện thoại cơ quan: 7.563.474, số máy lẻ 211.

MỤC LỤC

Chương 1
GIẢI CÁC BÀI TOÁN QUY HOẠCH TOÁN HỌC TRÊN MICROSOFT EXCEL

Dùng Solver ta có thể tìm cực đại hay cực tiểu của một hàm số đặt trong một ô gọi là ô đích. Solver chỉnh sửa một nhóm các ô (gọi là các ô có thể chỉnh sửa) có liên quan trực tiếp hay gián tiếp đến công thức nằm trong ô đích để tạo ra kết quả. Ta có thể thêm vào các ràng buộc để hạn chế các giá trị mà Solver có thể dùng. Đối với bài toán quy hoạch tuyến tính Solver dùng phương pháp đơn hình, đối với quy hoạch phi tuyến Solver dùng phương pháp tụt gradient để tìm một cực trị địa phương.

1.1. BÀI TOÁN QUY HOẠCH TUYẾN TÍNH CÓ MỘT CHỈ SỐ

Xét bài toán quy hoạch

$$c_1x_1 + c_2x_2 + \cdots + c_nx_n \quad = \quad f(x) \to \max / \min \tag{1}$$

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \quad Q \quad b_1$$

$$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \quad Q \quad b_2$$

$$\cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots$$

$$a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \quad Q \quad b_m$$

$$x_j \begin{cases} \geq 0 \\ = \text{interger} \\ = \text{binary (0 or 1)} \end{cases} \qquad j = 1, \ldots, n$$

trong đó Q là một trong các phép toán quan hệ $\geq \ \leq \ =$, thứ tự các phép toán quan hệ trong các ràng buộc là tuỳ ý. Như vậy bài toán (1) có thể là bài toán quy hoạch tuyến tính thông thường, quy hoạch tuyến tính nguyên hay quy hoạch boolean.

Cách bố trí dữ liệu cho trên bảng tính:

c[1]	c[2]	. . . . . .	c[n]	∑ c[j] x[j]	
a[1,1]	a[1,2]	. . . . . .	a[1,n]	∑ a[1,j] x[j]	b[1]
a[2,1]	a[2,2]	. . . . . .	a[2,n]	∑ a[2,j] x[j]	b[2]
. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .
a[m,1]	a[m,2]	. . . . . .	a[m,n]	∑ a[m,j] x[j]	b[m]
x[1]	x[2]	. . . . . .	x[n]		

Hàng cuối cùng là các giá trị ban đầu của các biến để các công thức của Excel hoạt động, có thể lấy giá trị của tất cả các biến bằng 1.

Xét bài toán:

$$x_1 + 4x_2 + x_3 \quad \rightarrow \quad \min \qquad (2)$$

$$2x_1 + 3x_2 + 4x_3 \geq 20$$

$$5x_1 - x_2 + 2x_3 \geq 12$$

$$x_1 + 2x_2 - x_3 \leq 2$$

$$-x_1 + 4x_2 - 2x_3 \leq 1$$

$$x_1, \quad x_2, \quad x_3 \geq 0$$

Các bước thực hiện để giải bài toán:

Bước 1. Nhập dữ liệu bài toán vào bảng tính dưới dạng sau:

	A	B	C	D	E	F
1		Bien 1	Bien 2	Bien 3	Gia tri ve trai	
2	Ham muc tieu	1	4	1		
3	Rang buoc 1	2	3	4		20
4	Rang buoc 2	5	-1	2		12
5	Rang buoc 3	1	2	-1		2
6	Rang buoc 4	-1	4	-2		1
7	Phuong an	1	1	1		
8						

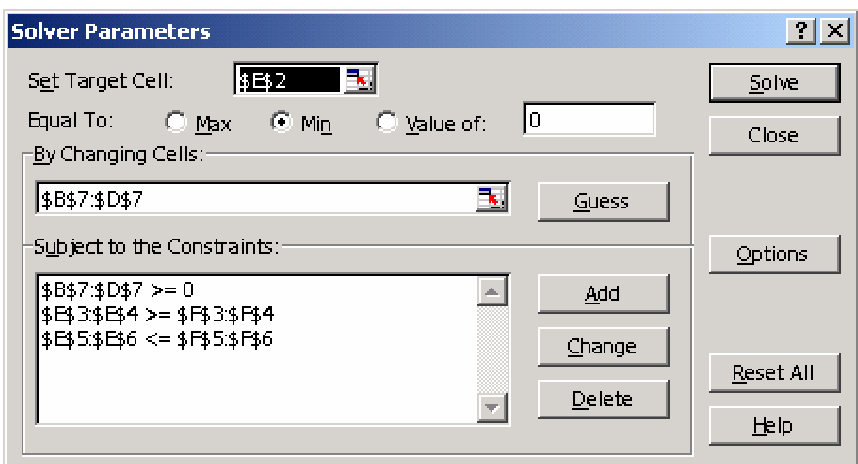
Phương án ban đầu X = (1, 1, 1), nó có thể không chấp nhận được.

Bước 2. Tính giá trị hàm mục tiêu tại ô E2 bằng công thức

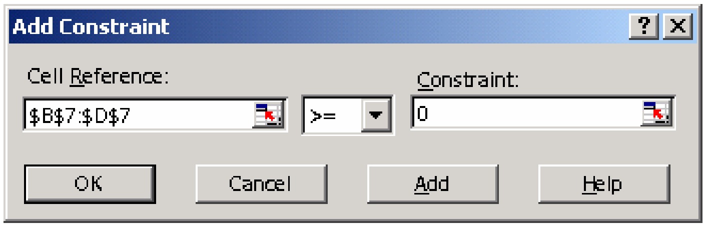
= SUMPRODOCT(B7 : D7, B2 : D2)

Hàm Sumproduct cho tích vô hướng của hai dãy ô. Copy công thức từ ô E2 sang dãy các ô E3 : E6 nhằm tính giá trị vế trái của bốn ràng buộc bài toán (1).

Bước 3. Dùng lệnh Tools / Solver, xuất hiện hộp thoại Solver Parameters.



Mục ***Set Target Cell***: chọn ô đích (chứa giá trị hàm mục tiêu), có thể nháy vào biểu tượng của Excel bên phải hộp văn bản để xác định ô, trong ví dụ chọn ô E2. Mục ***Equal To***: chọn Max nếu cực đại hàm mục tiêu, chọn Min nếu cực tiểu hàm mục tiêu, chọn Value of và nhập giá trị nếu muốn ô đích bằng một giá trị nhất định, trong ví dụ chọn Min. Mục By Changing cells: chọn các ô chứa các biến của bài toán, ta chọn khối ô B7:D7. Nháy nút Add để nhập tất cả các ràng buộc vào khung Subject to the Constraints (dòng đầu trong khung ứng với ràng buộc không âm trên các biến, dòng thứ hai ứng với hai ràng buộc đầu bài toán (2), dòng cuối ứng với 2 ràng buộc cuối). Khi nháy nút Add, hiện hộp thoại



Hộp văn bản Cell Reference để chọn các ô cần đặt ràng buộc lên chúng, hộp văn bản ở giữa để chọn loại ràng buộc (>= = <= interger, binary), hộp văn bản Constraint để chọn giá trị ràng buộc (có thể là số hay giá trị trong các ô).

Sau khi nhập xong các ràng buộc, nháy vào nút Options, hiện hộp thoại Solver Options, đánh dấu kiểm vào mục Assume Linear Model (khẳng định mô hình của ta là tuyến tính).

Bước 4. Trong hộp thoại Solver Parameters nháy vào nút Solve để bát đầu giải bài toán tối ưu. Giải xong bài toán xuất hiện hộp thoại Solver Results, chọn mục Keep Solver Solution (giữ lại lời giải), nháy OK, kết quả giải bài toán nằm ở các ô B7 : D7. Kết quả ta được phương án tối ưu là X = (0.5 , 0 , 4.75), giá trị cực tiểu hàm mục tiêu là 5.25 ở ô E2.

	A	B	C	D	E	F
1		Bien 1	Bien 2	Bien 3	Gia tri ve trai	
2	Ham muc tieu	1	4	1	5.25	
3	Rang buoc 1	2	3	4	20	20
4	Rang buoc 2	5	-1	2	12	12
5	Rang buoc 3	1	2	-1	-4.25	2
6	Rang buoc 4	-1	4	-2	-10	1
7	Phuong an	0.5	0	4.75		

1.2. BÀI TOÁN QUY HOẠCH TUYẾN TÍNH CÓ HAI CHỈ SỐ

Bài toán vận tải. Có m kho hàng (điểm phát) chứa một loại hàng hoá, lượng hàng ở kho i là a_i. Có n nơi tiêu thụ (điểm thu) loại hàng này, nhu cầu nơi j là b_j. Chi phí vận chuyển một đơn vị hàng từ điểm phát i tới điểm thu j là c_{ij}. Xác định các lượng hàng vận chuyển x_{ij} từ các điểm phát i tới các điểm thu j sao cho tổng chi phí là nhỏ nhất và nhu cầu các điểm thu được thoả mãn. Dạng toán học của bài toán:

$$\sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} \to \min \qquad (3)$$

$$\sum_{j=1}^{n} x_{ij} \le a_i \quad i = 1,\cdots,m$$

$$\sum_{i=1}^{m} x_{ij} \ge b_j \quad j = 1,\cdots,n$$

$$x_{ij} \ge 0 \quad i = 1,\cdots m, \quad j = 1,\cdots,n$$

	Điểm thu 1	Điểm thu 2		Điểm thu n	Trị mục tiêu
Điểm phát 1	c[1,1]	c[1,2]	. . . . . .	c[1,n]	∑ c[i,j] x[i,j]
Điểm phát 2	c[2,1]	c[2,2]	. . . . . .	c[2,n]	
Điểm phát 3	. . . . . .	. . . . . .	. . . . . .	. . . . . .	
Điểm phát 4	c[m,1]	c[m,2]	. . . . . .	c[m,n]	

					Cộng hàng	Khả năng
Phương án	x[1,1]	x[1,2]	. . . . . .	x[1,n]	∑ x[1,j]	a[1]
	x[2,1]	x[2,2]	. . . . . .	x[2,n]	∑ x[2,j]	a[2]
	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .
	x[m,1]	x[m,2]	. . . . . .	x[m,n]	∑ x[m,j]	a[m]
Cộng cột	∑ x[i,1]	∑ x[i,2]	. . . . . .	∑ x[i,n]		
Nhu cầu	b[1]	b[2]	. . . . . .	b[n]		

Ví dụ. Xét bài toán vận tải có 3 điểm phát và 4 điểm thu được nhập vào bảng tính:

	A	B	C	D	E	F
1		Ma tran chi phi				
2	2	1	4	3		Ham muc tieu
3	6	0	5	2		38
4	1	4	8	2		
5						
6		Phuong an van chuyen			Tong hang	
7	1	1	1	1	4	10
8	1	1	1	1	4	25
9	1	1	1	1	4	15
10	3	3	3	3	Tong cot	
11	5	15	20	10		

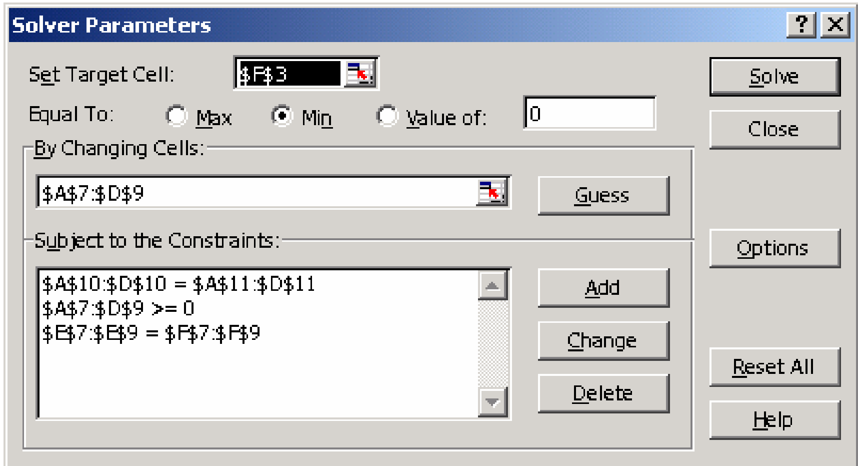
Khối A2:D4 là ma trận chi phí vận chuyển, khối A7:D9 là phương án vận chuyển (giá trị ban đầu cho tất cả bằng 1), khối F7:F9 là khả năng của 3 điểm phát, khối A11:D11 là nhu cầu của 4 điểm thu, khối E7:E9 là lượng hàng phát từ mỗi điểm phát i theo phương án X đã chọn, khối A10:D10 là lượng hàng nhận được tại mỗi điểm thu j theo phương án X. Giả sử rằng các bất đẳng thức trong dòng thứ hai và thứ ba của bài toán (3) là đẳng thức và tổng lượng hàng có trong các kho bằng tổng nhu cầu của các nơi thiêu thụ.

Quá trình dùng Solver để giải bài toán vận tải trên theo các bước:

Bước 1. Nhập chi phí vận chuyển vào các ô A2:D4, nhập khả năng của các điểm phát vào F7:F9, nhu cầu các điểm thu A11:D11, phương án ban đầu A7:D9.

Bước 2. Tính giá trị hàm mục tiêu trong ô F3 theo công thức = Sumproduct (A2:D4, A7:D9), hàm này tính tổng các tích của từng cặp phần tử trong hai khối ô. Tính lượng hàng phát của điểm phát 1 tại ô E7 theo công thức =SUM(A7:D7), sao chép công thức này vào các ô E8:E9. Tính lượng hàng nhận được của điểm thu 1 tại ô A10 theo công thức = SUM(A7:A9), sao chép công thức này vào các ô B10:D10.

Bước 3. Dùng lệnh Tools/ Solver với các lựa chọn hàm mục tiêu và các ràng buộc:



Trong hộp thoại Solver Options phải chọn Assume Linear Model. Cuối cùng ta nhận được giá trị tối ưu hàm mục tiêu bằng 115, phương án vận chuyển tối ưu: x[1,3]= 10, x[2,2]= 15, x[2,3]= 10, x[3,1]= 5, x[3,4]= 10 trong bảng tính kết quả:

	A	B	C	D	E	F
1		Ma tran chi phi				
2	2	1	4	3		Ham muc tieu
3	6	0	5	2		115
4	1	4	8	2		
5						
6		Phuong an van chuyen			Tong hang	
7	0	0	10	0	10	10
8	0	15	10	0	25	25
9	5	0	0	10	15	15
10	5	15	20	10	Tong cot	
11	5	15	20	10		

1.3. GIẢI BÀI TOÁN QUY HOẠCH PHI TUYẾN

Xét bài toán quy hoạch phi tuyến

$$Min\left\{f(x)\ \middle|\ g_i(x)=0,\quad i=1,2,\cdots,m,\quad x\in R^n\right\}$$

Để giải bài toán quy hoạch phi tuyến bằng Solver ta cần xác định khối ô để chứa các biến (x[1], x[2], . . . , x[n]), một ô chứa giá trị hàm mục tiêu f(x), khối m ô chứa giá trị các hàm $g_i(x)$.

Ví dụ giải bài toán quy hoạch toàn phương:

$$-x_1-2x_2+0.5x_1^2+0.5x_2^2 \quad \to Min$$
$$2x_1+3x_2+x_3=6$$
$$x_1+4x_2+x_4=5$$
$$x_1,x_2,x_3,x_4\ge 0$$

Bảng tính để giải bài toán này như sau:

	A	B	C	D	E	F	G
1		X1	X2	X3	X4	f(x)	
2	Phuong an	0	0	0	0	0	
3	Rang buoc 1	2	3	1	0	0	6
4	Rang buoc 2	1	4	0	1	0	5

Phương án trong khối ô B2:E2 (phương án ban đầu cho mọi phần tử bằng 0), hàm mục tiêu trong ô F2 xác định bởi công thức = - b2 - 2*c2 + 0.5*b2^2 + 0.5*c2^2. Ô F3 tính theo công thức = sumproduct (b2: e2, b3 : e3), công thức này chép sang ô F4. Các ràng buộc bài toán B2 : E2 >= 0, và F3:F4 = G3:G4. Khi giải các bài toán quy hoạch phi tuyến ta phải bỏ chọn mục Assume Linear Model trong hộp thoại Solver Options. Kết quả dùng Solver giải bài toán: trị tối ưu -2.0294, phương án tối ưu (0.7647, 1.0588, 1.294, 0).

Tóm lại Solver có thể giải được nhiều bài toán tối ưu, số lượng biến tối đa của bài toán là 200 biến. Tuy nhiên cũng có nhiều bài toán nó không giải được, khi đó nó đưa thường đưa ra các thông báo:

– Solver could not find a feasible solution: bài toán không có lời giải chấp nhận được. Hoặc có thể do các giá trị khởi đầu của những ô chứa biến khác quá xa các giá trị tối ưu, hãy thay đổi các giá trị khởi đầu và giải lại.

– The maximum iteration limit was reached, continue anyway ? số bước lặp đã đến số cực đại. Ta có thể tăng số bước lặp ngầm định nhờ lệnh Tools/ Solver, chọn Options, nhập giá trị mới vào hộp Iterations.

– The maximum time limit was reached, continue anyway ? thời gian chạy vượt quá thời gian tối đa ngầm định. Ta có thể sửa giá trị trong mục Max Time trong gộp thoại Solver Options.

Chú ý, nếu các lệnh Solver và Data Analysis không có trong menu Tools ta phải cài đặt bổ sung từ đĩa CD: dùng lệnh Tools / Add-Ins, hiện hộp thoại, chọn mục Solver Add in và Analysis ToolPak.

BÀI TẬP

1. Giải bài toán ***quy hoạch tuyến tính nguyên bộ phận***:

$$z = 5x_1 + x_2 + x_3 + 2x_4 + 3x_5 \to \min$$
$$-x_2 + 5x_3 - x_4 - 2x_5 \le 2$$
$$5x_1 - x_2 + x_5 \ge 7$$
$$x_1 + x_2 + 6x_3 + x_4 \ge 4$$
$$x_j \ge 0 \quad j = 1, 2, 3, 4, 5$$
$$x_j = \text{interger}, \quad j = 1, 2, 3$$

Đáp số: trị tối ưu là 12, phương án tối ưu (2, 2, 0, 0, 0).

2. Giải bài toán ***quy hoạch tuyến tính boolean*** (bài toán cái túi) sau:

$$30x_1 + 19x_2 + 13x_3 + 38x_4 + 20x_5 + 6x_6 + 8x_7 + 19x_8 + 10x_9 + 11x_{10} \to \max$$
$$15x_1 + 12x_2 + 9x_3 + 27x_4 + 15x_5 + 5x_6 + 8x_7 + 20x_8 + 12x_9 + 15x_{10} \le 62$$
$$x_j \in \{0, 1\}, \quad j = 1, 2, \cdots, 10$$

Đáp số: trị tối ưu là 95, phương án tối ưu là (1, 1, 0, 1, 0, 0, 1, 0, 0, 0).

3. ***Giải bài toán phân công công việc***. Có n đơn vị sản xuất cần sản xuất n loại sản phẩm, đơn vị i sản xuất sản phẩm j với chi phí là c[i,j]. Hãy phân công mỗi đơn vị sản xuất một sản phẩm để tổng chi phí là nhỏ nhất. Dạng bài toán:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} x_{ij} \to \min$$
$$\sum_{j=1}^{n} x_{ij} = 1, \quad i = 1, 2, \cdots, n$$

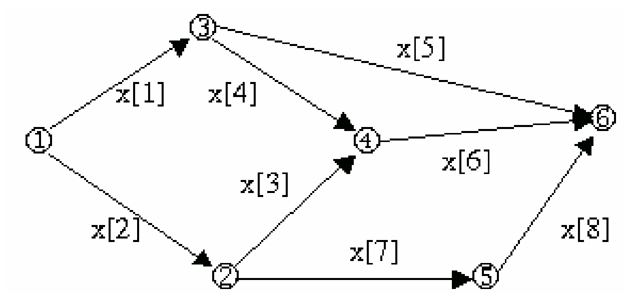
$$\sum_{i=1}^{n} x_{ij} = 1, \quad j = 1, 2, \cdots, n, \qquad x_{ij} = \{0, 1\}$$

Dùng Solver giải bài toán phân công với n = 4 và ma trận chi phí sau:

100000	4000000	800000	550000
200000	3500000	750000	500000
400000	2000000	700000	400000
300000	5000000	600000	450000

Đáp số: trị tối ưu hàm mục tiêu là 3200000 đồng, phương án tối ưu là x[1,1]= x[2,4]= x[3,2]= x[4,3] = 1.

4. Giải bài toán ***tìm luồng cực đại trên đồ thị có hướng***. Cho đồ thị có hướng gồm 6 đỉnh, nếu từ đỉnh u tới đỉnh đỉnh v có đường vận chuyển thì ta vẽ một cung j, lượng hàng vận chuyển trên cung này là x[j], khả năng vận chuyển tối đa trên cung này là q[j]. Tìm lượng hàng lớn nhất có thể vận chuyển từ đỉnh 1 đến đỉnh 6.



Từ đồ thị trên ta có thể viết hàm mục tiêu và các ràng buộc như sau:

x[1] + x[2] ===> Max
x[1] - x[4] -x[5] = 0
x[2] - x[3] - x[7] = 0
x[3] + x[4] - x[6] = 0
x[7] - x[8] = 0
x[1] + x[2] - x[5] - x[6] - x[8] = 0
0 <= x[j] <= q[j], j = 1, 2, 3, 4, 5, 6, 7, 8

trong đó véc tơ q = (4, 2, 4, 4, 1, 2, 2, 2).

Đáp số: lượng hàng tối đa có thể vận chuyển là 5, phương án tối ưu là x = (3, 2, 0, 2, 1, 2, 2, 2). Đối với các bài toán đồ thị khác, nếu bằng cách đặt biến ta có thể phát biểu chúng dưới dạng đại số thì cũng có thể dùng Solver để giải.

5. Giải bài toán ***quy hoạch lõm*** (có thể có nhiều cực tiểu địa phương)

$$-x_1^2 + 2x_1 - x_2^2 + 4x_2 - x_3^2 + 8x_3 - x_4^2 + 14x_4 - x_5^2 + 18x_5 - 180 \rightarrow Min$$

$$-x_1 - 2x_2 + x_3 + 2x_4 + 3x_5 \le 85$$

$$-7x_1 + 9x_2 - 5x_3 + 33x_4 - 11x_5 \le 500$$

$$2x_1 - x_2 + 2x_3 - x_4 + 2x_5 \le 150$$

$$1.3x_1 + x_2 + x_3 + x_4 + x_5 \le 300$$

$$x_1 + x_2 + x_3 + x_4 + x_5 \le 300$$

$$x_1, x_2, x_3, x_4, x_5 \ge 0$$

Với phương án ban đầu X = (50, 50, 50, 50, 50) dùng Solver ta được lời giải tối ưu là X = (0, 190, 0, 0, 110) và trị tối ưu hàm mục tiêu là - 45640.

Chương 2
GIẢI CÁC BÀI TOÁN THỐNG KÊ TRÊN MICROSOFT EXCEL

Trong Excel có một số hàm mảng để thực hiện hồi quy tuyến tính (Linest, Trend, Forecast, Slope, Intercept) và hồi quy mũ (Logest, Growth). Những hàm này được nhập như những công thức mảng và cho kết quả mảng.

2.1. Hồi quy tuyến tính bội

Phương trình hồi quy bội tuyến tính có dạng:

$$y = m_1 x_1 + m_2 x_2 + \cdots + m_n x_n + b, \qquad (1)$$

trong đó x1, x2, . . . , xn là các biến độc lập, y là biến phụ thuộc, các hệ số m1, m2, …, mn, b là các hệ số cần xác định. Các giá trị quan sát của các biến có thể bố trí theo dạng cột hoặc theo dạng hàng.

• ***Hàm LINEST*** dùng để tính các hệ số của phương trình hồi quy tuyến tính bội, cú pháp:

= LINEST(known_y's, known_x's, const, stats)

trong đó ***known_y's*** là khối ô chứa các quan sát của biến y; ***known_x's*** là khối ô chứa các quan sát của các biến x1, x2, . . . , xn; biến ***const*** có giá trị logic (nhập *True hoặc để trống* nếu có tính b, nhập False nếu buộc b=0). Biến ***stats*** có giá trị logic, nhập False nếu không in các thống kê hồi quy, nhập *True hoặc bỏ trống* thì hàm cho các thống kê hồi quy dạng:

$$\begin{array}{llllll} m_n & m_{n-1} \cdots\cdots\cdots m_2 & m_1 & b \\ se_n & se_{n-1} \cdots\cdots\cdots & se_2 & se_1 & se_b \\ r^2 & se_y \\ F & d_f \\ ss_{reg} & ss_{resid} \end{array}$$

trong đó $se_n \quad se_{n-1} \cdots\cdots\cdots \quad se_2 \quad se_1 \quad se_b$ là các sai số chuẩn hoá của các hệ số m1, m2, ..., mn, b. Hệ số r2 là hệ số xác định thuộc [0, 1], nếu r2 = 1 thì có quan hệ hoàn hảo trong mẫu, nếu r2 = 0 thì phương trình hồi quy không có tác dụng dự đoán y.

Hệ số se_y là sai số chuẩn hoá cho ước lượng y. Hệ số F là thống kê F, dùng F để xác định liệu giữa biến phụ thuộc và các biến độc lập có thực sự quan hệ với nhau hay đó chỉ là thể hiện của tác động ngẫu nhiên. Hệ số d_f là bậc tự do, dùng để xác định mức tin cậy của mô hình hồi quy. Các hệ số $ss_{reg} \quad ss_{resid}$ là tổng bình phương giá trị hồi quy và tổng bình phương độ lệch.

Ví dụ 1. Bảng bên cho số liệu về doanh thu (Y), chi phí cho quảng cáo (X1), tiền lương của nhân viên tiếp thị (X2) của 12 công ty tư nhân, đơn vị là 1 triệu đồng. Xây dựng hàm hồi quy tuyến tính bội Y phụ thuộc vào X1, X2.

	A	B	C	D
1	Y	X1	X2	Trend
2	127	18	10	124.9673
3	149	25	11	147.2661
4	106	19	6	108.4383
5	163	24	16	168.5539
6	102	15	7	103.1741
7	180	26	17	178.324
8	161	25	14	161.5422
9	128	16	12	129.4732
10	139	17	12	131.979
11	144	23	12	147.0134
12	159	22	14	154.025
13	138	15	15	141.2436

Để ước lượng hàm hồi quy ta dùng hàm mảng Linest như sau: đánh dấu khối vùng ô B19: D23, nhập công thức =LINEST(A2 : A13, B2 : C13, True, True), ấn Ctrl + Shift + Enter, kết quả ta được 12 số:

	A	B	C	D
18		m2	m1	b
19		4.75869	2.505729	32.27726
20	SE	0.41038	0.328573	6.253073
21	R^2	0.97566	4.003151	#N/A
22	F	180.355	9	#N/A
23	SS	5780.44	144.2269	#N/A

Như vậy phương trình hồi quy là

$$Y = 2.505729\ X1 + 4.75869\ X2 + 32.27726. \qquad (2)$$

• ***Hàm TREND*** nhằm tính các giá trị y theo hàm ước lượng (1) với các bộ giá trị cho trước (x1, x2, . . . , xn), các bộ giá trị này có thể là các quan sát cũ hoặc các dự báo mới. Cú pháp hàm:

= TREND(known_y's, known_x's, new_x's, const)

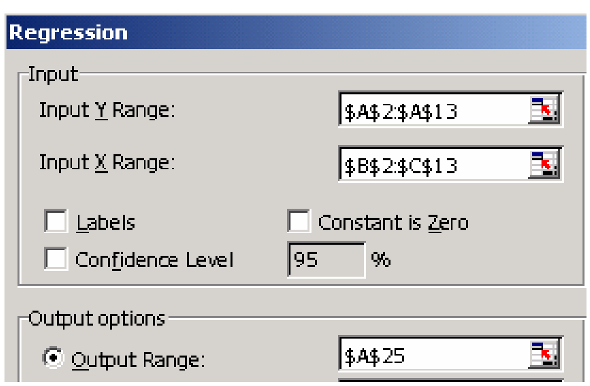
trong đó ***know_y's*** là khối ô chứa chứa các quan sát của biến y; ***known_x's*** là khối ô chứa các quan sát của các biến x1, x2, . . . , xn; biến ***const*** có giá trị logic (nhập *True hoặc để trống* nếu có tính b, nhập False nếu buộc b=0). Tham số ***new_x's*** là khối ô chứa các giá trị mới của x1, x2, . . . , xn mà ta cần tính các giá trị y tương ứng theo (1); nếu bỏ trống tham số này thì coi nó chính là know_x's.

Trở lại ví dụ 1, dùng hàm Trend tính cột D (là các giá trị y tính theo (2) với các bộ giá trị x1, x2, …, xn tương ứng trong khối B2 : C13). Thao tác tính: đánh dấu khối vùng ô chứa kết quả của hàm là D2 : D13, nhập công thức = Trend(a2:a13, b2:c13), ấn Ctrl + Shift + Enter. So sánh khối ô D2:D13 với khối ô A2:A13 ta thấy được sự sai khác giữa giá trị y tính theo hàm (2) với giá trị thực tế quan sát được.

Tiếp theo, cho các bộ giá trị mới x1, x2 trong khối ô B15 : C17, cần dự báo các giá trị y được tính theo (2) trong khối ô D15 : D17. Thao tác tính: đánh dấu khối vùng ô D15:D17, nhập công thức = Trend(a2: a13, b2: c13, b15: c17, True), ấn Ctrl + Shift + Enter.

	A	B	C	D
14				Du bao
15		26	18	183.0827
16		28	19	192.8529
17		30	20	202.623

• ***Lệnh Tools / Data Analysis*** nhằm tính các tham số của hàm hồi quy tuyến tính bội (1) và các thống kê. Xét ví dụ 1, giả sử ta đã nhập các quan sát của các biến y, x1, x2 trong khối ô A2: C13. Dùng lệnh Tools / Data Analysis, hiện hộp thoại Data Analysis, chọn mục Regression, nháy OK, hiện hộp thoại Regression:



nhập khối ô chứa các quan sát của biến y vào mục Input Y Range, nhập khối ô chứa các giá trị quan sát của biến x1, x2 vào mục Input X Range, lựa chọn mục Output Range và nhập địa chỉ ô ở góc trên bên trái của vùng chứa kết quả, nháy OK. Kết quả cho trong bảng sau:

	A	B	C	D	E	F
26	*Regression Statistics*					
27	Multiple R	0.9878				
28	R Square	0.9757				
29	Adjusted R Square	0.9702				
30	Standard Error	4.0032				
31	Observations	12.0000				
32		*df*	*SS*	*MS*	*F*	*Significance F*
33	Regression	2	5780.4397	2890.2199	180.3545	0.0000
34	Residual	9	144.2269	16.0252		
35	Total	11	5924.6667			
36		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*
37	Intercept	32.2773	6.2531	5.1618	0.0006	18.1318
38	X Variable 1	2.5057	0.3286	7.6261	0.0000	1.7624
39	X Variable 2	4.7587	0.4104	11.5957	0.0000	3.8303

Trong bảng trên Hệ số xác định r2 nằm trong ô B28, sai số chuẩn hoá cho ước lượng y nằm trong ô B30, khối ô B37: B39 chứa các hệ số đường hồi quy b, m1, m2. Khối ô C37: C39 chứa các sai số chuẩn hoá của b, m1, m2. Thống kê F trong ô E33.

2.2. Hồi quy tuyến tính đơn

Hồi quy tuyến tính đơn là trường hợp riêng của hồi quy tuyến tính bội (1) với n=1:

$$y = mx + b \qquad (3)$$

Do đó tất cả các hàm và lệnh đã trình bày với hồi quy tuyến tính bội cũng đúng với hồi quy tuyến tính đơn. Song đối với hồi quy tuyến tính đơn có thêm ba hàm mới.

– ***Hàm Slope(known_y's, known_x's)*** ước lượng giá trị m của phương trình (3).

– ***Hàm Intercept(known_y's, known_x's)*** ước lượng giá trị b của (3).

– ***Hàm Forecast(x, known_y's, known_x's)***: dự đoán y theo phương trình (3) với giá trị x biết trước.

Ví dụ 2. Tính hàm hồi quy của y (sản lượng nông nghiệp) phụ thuộc vào x (lượng phân bón).

	A	B	C	D	E
1	Y	X		m	b
2	13983.800	763.534		4.064894	11456.13
3	14406.400	784.630			
4	15005.300	776.200		X	Du bao Y
5	16829.000	1118.600		1612	18008.740
6	17100.000	1488.000			

Công thức trong ô D2 là = Slope(a2:a6, b2:b6), công thức trong ô E2 là =Intercept(a2:a6, b2:b6), công thức trong ô E5 là =Forecast(d5, a2:a6, b2:b6) để dự báo y với x = 1612.

2.3. Hồi quy mũ

Phương trình hồi quy mũ là

$$y = b * m_1^{x_1} * m_2^{x_2} * \cdots * m_n^{x_n} \qquad (4)$$

Nếu chỉ có một biến độc lập phương trình sẽ là $y = b * m^x$.

Hàm Logest dùng để ước lượng các hệ số của phương trình (4), nó làm việc giống như hàm Linest (các đối số và mảng kết quả hoàn toàn giống). Cú pháp:

= LOGEST(known_y's, known_x's, const, stats).

Hàm Growth dùng để tính các giá trị y theo (4) với các bộ giá trị (x1, x2, … , xn) cho trước, làm việc hoàn toàn giống hàm Trend. Cú pháp:

= GROWTH(known_y's, known_x's, new_x's, const).

BÀI TẬP

1. Cho Y là nhu cầu thịt bò (đơn vị 100 tấn) của 12 tháng liên tiếp (X) trong một khu dân cư:

X: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

Y: 15, 18, 18, 16, 14, 18, 20, 21, 19, 20, 24, 26.

Hãy ước lượng hàm hồi quy tuyến tính đơn, dự báo nhu cầu thịt bò cho 3 tháng tiếp theo. Đáp số : y = 0.793706 x + 13.92424.

2. Trong 10 tháng liên tiếp lượng hàng bán ra của một công ty rất thấp, sau đó công ty tung ra thị trường một sản phẩm mới và nhận thấy lượng hàng bán ra tăng theo hàm mũ. Số đơn vị hàng bán ra (Y) trong 6 tháng tiếp theo (X) cho trong bảng sau:

	A	B
1	Y	X
2	33100	11
3	47300	12
4	69000	13
5	102000	14
6	150000	15
7	220000	16

Hãy ước lượng hàm hồi quy mũ và dự báo lượng hàng bán ra trong các tháng 17, 18, 19, 20 (dùng hàm Growth). Đáp số : $y = 495.3048 * 1.463276^{x}$.

3. Tính hàm hồi quy tuyến tính bội với số liệu cho trong bảng dưới

	A	B	C	D	E
1	Y	X1	X2	X3	X4
2	733.300	3.089	76.200	283.500	15.844
3	750.900	3.503	79.400	274.500	19.835
4	747.600	3.817	77.000	268.000	21.797
5	727.600	3.870	74.000	265.700	24.759
6	694.400	3.706	64.400	259.600	28.093
7	702.600	3.851	63.100	256.800	31.121
8	714.000	4.170	66.300	259.300	32.759
9	717.630	4.378	62.900	263.400	34.556
10	750.000	5.000	66.700	273.100	36.788

trong đó Y là thu nhập quốc dân, X1 là sản lượng điện, X2 là sản lượng than, X3 là sản lượng lương thực, X4 là sản lượng thép. Dùng hai phương pháp: dùng hàm Linest và lệnh Tools / Data Analysis. Dự báo Y với X = (5.2, 65.1, 275.3, 37.8). Đáp số: dự báo Y = 751.79289.