

# NG D NG EXCEL TRONG H I QUY T NG QUAN VÀ D BÁO KINH T

Các y u t trong m i ho t ng s n xu t kinh doanh luôn có m i liên h m t thi t v i nhau. Xác nh tính ch t ch t ch c a các m i liên h gi a các y u t và s d ng các s li u ã bi t đ báo s giúp nhà qu n lý r t nhi u trong vi c ho ch nh các k ho ch s n xu t kinh doanh hi n t i c ng nh trong t ng lai.

## 4.1 NG D NG EXCEL TRONG D BÁO KINH T

### 4.1.1 Ý ngh a c a đ báo kinh t

*D báo* là phán đoán nh ng s ki n s x y ra trong t ng lai trên c s phân tích khoa h c các đ li u c a quá kh và hi n t i nh m t s mô hình toán h c. *D báo kinh t* là vi c a ra các đ báo nh ng s ki n kinh t s x y ra trong t ng lai đ a trên c s phân tích khoa h c các s li u kinh t c a quá kh và hi n t i. Ch ng h n, nhà qu n lý đ a trên c s các s li u v doanh thu bán hàng c a k tr c và k này a ra đ báo v th tr ng ti m n ng c a doanh nghi p trong t ng lai. Do ó, trong ho t ng s n xu t kinh doanh đ báo em l i ý ngh a r t l n. Nó là c s l p các k ho ch qu n tr s n xu t và marketing t o tính hi u qu và s c c nh tranh cho các chi n l c s n xu t trong t ng lai.

D báo mang tính khoa h c và òi h i c m t ngh thu t đ a trên c s phân tích khoa h c các s li u thu th p c. B i l c ng đ a vào các s li u th i gian nh ng l y s l ng là bao nhiêu, m c nh ng th i gian cu i nhi u hay ít s khi n cho mô hình đ oán ph n ánh y hay không y nh ng thay i c a các nhân t m i i v i s bi n ng c a hi n t ng. Do v y mà đ báo v a mang tính ch quan v a mang tính khách quan. D báo mu n chính xác thì càng c n ph i lo i tr tính ch quan c a ng i đ báo.

### 4.1.2 Gi i thi u các ph ng pháp đ báo kinh t

Ngày nay d báo ã c s d ngr t r ng rãi trong m i l nh v c c a i s ng xã h i v i nhi u lo i và ph ng pháp d báo khác nhau nh ph ng pháp l y ý ki n ban i u hành, ph ng pháp i u tra ng i tiêu dùng, ph ng pháp Delphi... Trong th ng kê ng i ta s d ngr t nhi u ph ng pháp khác nhau nh : ph ng pháp trung bình gi n n, ph ng pháp trung bình dài h n, ph ng pháp san b ng hàm m ... Ch ng này c p n ba ph ng pháp là: ph ng pháp trung bình dài h n, ph ng pháp trung bình ng, ph ng pháp h i quy t ng quan...

➤ **Ph ng pháp trung bình dài h n:** S d báo b ng trung bình c ng c a các quan sát th c t tr c ó.

$$\text{Công th c: } F_{t+1} = \frac{\sum_{i=0}^{n-1} D_{t-i}}{n}$$

Trong ó:  $F_{t+1}$  là s d báo k th  $t+1$

$D_t$  là s quan sát k th  $t$

$n$  t ng s quan sát

Ph ng pháp này làm san b ng s ng u nhiên, nó phù h p v i nh ng mô hình mà các l ng t ng (gi m) tuy t i liên hoàn x p x nhau (dòng yêu c u u). Tuy nhiên, kh i l ng tính toán nhi u và ph i l u tr nhi u s li u.

➤ **Ph ng pháp trung bình ng:** S d báo k th  $t+1$  b ng trung bình c ng c a n k tr c ó. Nh v y, c m i k d báo l i b i s li u xa nh t trong quá kh và thêm vào s li u m i nh t.

$$\text{Công th c: } F_{t+1} = \frac{D_t + D_{t-1} + \dots + D_{t-n}}{n+1}$$

Th ng thì ng i ta l y n là khá nh  $n = 3, 4, 5 \dots$

ây c ng là ph ng pháp d báo phù h p v i nh ng mô hình mà các l ng t ng (gi m) tuy t i liên hoàn x p x nhau.

➤ **Phương pháp hồi quy tuyến tính**

Phân tích hồi quy nghiên cứu mối phụ thuộc của một biến (gọi là biến phụ thuộc hay biến độc lập) với một hay nhiều biến khác (gọi là (các) biến độc lập hay biến giải thích có giá trị đã biết) nhằm xác định và dự báo giá trị trung bình của biến phụ thuộc với các giá trị đã biết của (các) biến độc lập.

+ **Mô hình hồi quy tuyến tính** (mô hình hồi quy tuyến tính): là mô hình hồi quy nói lên mối phụ thuộc tuyến tính của một biến phụ thuộc với một hay nhiều biến độc lập mà phương trình của mô hình hồi quy có dạng tuyến tính với các hệ số.

Mô hình hồi quy tuyến tính có dạng:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + U_i$$

Trong đó  $U_i$  là sai số ngẫu nhiên.

Mô hình hồi quy tuyến tính có dạng:

$$y = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_k X_{ki} + u_i$$

Trong đó,  $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \dots, \hat{\beta}_k$  là các hệ số ước lượng và không chắc chắn  $\beta_1, \beta_2, \beta_3, \dots, \beta_k$  bằng phương pháp bình phương nhỏ nhất. Nó là các sai số ngẫu nhiên, với mỗi  $u_i$  khác nhau chúng có giá trị khác nhau.  $u_i$  là các sai số ngẫu nhiên gây ra sai lệch của  $y$  với giá trị trung bình của nó.

Trong mô hình này ta chấp nhận giả thuyết các biến độc lập, không tác động và có phương sai không thay đổi. Trên thực tế, khi nghiên cứu các trường hợp thực tế thì ta tiến hành phân tích phương sai và tác động quan trọng nhất là dò dẫm của mối quan hệ phụ thuộc ngẫu nhiên khi kiểm tra xem có hiện tượng tác động quan trọng, các hệ số tuyến tính hay phương sai thay đổi không (thông dụng nhất là kiểm định Dolbin Watson).

Mô hình quan hệ tuyến tính trên đây xây dựng trên cơ sở mối liên hệ giữa một biến phụ thuộc  $Y$  và nhiều biến độc lập  $X$  gọi là **mô hình hồi quy tuyến tính bội**. Khi mô hình quan hệ tuyến tính xây dựng trên cơ sở mối liên hệ giữa hai biến (biến phụ thuộc  $Y$  và biến độc lập  $X$ ) thì gọi là **mô hình hồi quy tuyến tính đơn**.

Trên cơ sở thông tin thu được trong mẫu thống kê ta sẽ dùng phương pháp bình phương nhỏ nhất để tìm các hệ số của mô hình hồi quy. Tuy nhiên là dựa trên quan niệm rằng không chỉ giá trị quan sát của biến giải thích càng gần với giá trị thực của nó hay phần dư của chúng càng nhỏ càng tốt.

+ *Mô hình hồi quy phi tuyến*: là các dạng mô hình hồi quy phi tuyến nói lên mối phụ thuộc của một biến phụ thuộc với một hay nhiều biến độc lập mà phương trình của mô hình hồi quy có dạng phi tuyến với các hệ số. Chẳng hạn, như hàm sản xuất Cobb Douglas, hồi quy Parabol, hồi quy Hyperbol...

Như vậy, dựa vào các quan sát thu thập theo thời gian trong các kết cấu số liệu xây dựng mô hình hồi quy (cách xây dựng mô hình học trong môn Kinh tế lượng). Thay số liệu của các biến ã cho trong kết cấu báo vào mô hình hồi quy ta sẽ cho ta kết quả cần dựa báo.

### 4.1.3 Dự báo bằng phương pháp trung bình dài hạn trong Excel

#### *Quy trình dự báo:*

- Nhập số liệu thu thập được vào bảng tính.
- Sử dụng hàm *AVERAGE* tính ra số báo.

hiểu rõ hơn ta xét ví dụ minh họa sau:

*Ví dụ 4.1:* Một nhà phân ngành A ghi lại tỉ lệ hành thu thập sản phẩm trong 5 năm liên tiếp (2001-2005). Giả sử rằng tỉ lệ tăng trưởng sản phẩm hàng năm tăng dần như sau. Hãy dự báo sản phẩm sinh trong năm 2006 với số liệu như sau:

năm	2000	2001	2002	2003	2004	2005
sản phẩm sinh (bé)	29	30	28	31	29	26

#### *Bài tập:*

- Nhập số liệu thu thập được vào bảng tính theo đúng cột và sử dụng hàm *AVERAGE* dự báo ta có:

	A	B	C	D	E	F	G	H
6		<i>Phương pháp trung bình dài hạn dự báo số trẻ sơ sinh năm 2006</i>						
7								
8		<b>năm</b>	<b>số trẻ sơ sinh</b>	<b>dự báo</b>		<i>Công thức</i>		
9		2000	29	-		Nhập công thức cho ô D11		
10		2001	30	-		copy cho các ô D12:D15		
11		2002	28	30		D11=AVERAGE(\$C\$9:C10)		
12		2003	31	29		D12=AVERAGE(\$C\$9:C11)		
13		2004	29	30		D13=AVERAGE(\$C\$9:C12)		
14		2005	27	29		D14=AVERAGE(\$C\$9:C13)		
15		2006		<b>29</b>		D15=AVERAGE(\$C\$9:C14)		
16								
17		<i>Nhận xét:</i> Dự báo đến năm 2006 số trẻ sơ sinh của địa phương A sẽ là 29 bé						

Hình 4.1 Phương pháp dự báo trung bình dài hạn sử dụng hàm AVERAGE

#### 4.1.4 Dự báo bằng phương pháp trung bình động trong Excel

##### ➤ Quy trình dự báo bằng hàm AVERAGE

- Nhập số liệu thu thập được vào bảng tính.
- Sử dụng hàm **AVERAGE** tính ra số dự báo.

Xét ví dụ 4.1: Sử dụng phương pháp dự báo trung bình động 3 kỳ ta có:

	A	B	C	D	E	F	G	H
18		<i>Phương pháp trung bình động dự báo số trẻ sơ sinh năm 2006 với n=3</i>						
19								
20		<b>năm</b>	<b>số trẻ sơ sinh</b>	<b>dự báo</b>		<i>Công thức</i>		
21		2000	29	-		Nhập công thức cho ô D24		
22		2001	30	-		copy cho các ô D25:D27		
23		2002	28	-				
24		2003	31	29		D24=AVERAGE(C21:C23)		
25		2004	29	30		D25=AVERAGE(C22:C24)		
26		2005	27	29		D26=AVERAGE(C23:C25)		
27		2006		<b>29</b>		D27=AVERAGE(C23:C26)		
28								
29		<i>Nhận xét:</i> Dự báo đến năm 2006 số trẻ sơ sinh của địa phương A sẽ là 29 bé						

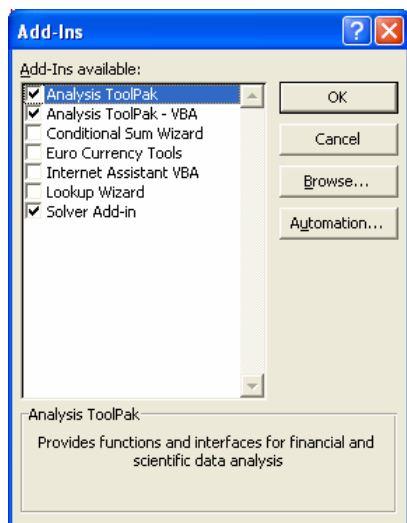
Hình 4.2 Phương pháp dự báo trung bình động sử dụng hàm AVERAGE

Ngoài quy trình dự báo sử dụng hàm **AVERAGE** trên ta còn có thể sử dụng trình cài thêm **Moving Average** không chỉ đưa ra giá trị dự báo mà còn đưa ra các sai số chuẩn và thể dự báo.

##### ❖ Bổ sung công cụ phân tích dữ liệu Data Analysis vào Excel

+ Khi dùng Excel

+ Vào thanh menu **Tools**, chọn **Add-Ins**. Hộp thoại **Add-Ins** xuất hiện tích vào mục **Analysis ToolPak** và **Analysis ToolPak VBA**.



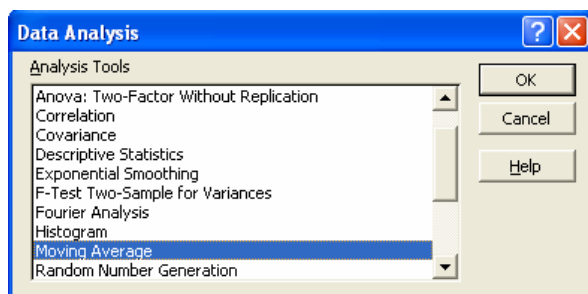
Hình 4.3 Bổ sung công cụ **Data Analysis**

+ Nhấn **OK** hoàn tất việc cài đặt. Lúc này trên thanh menu đã có một mục trong **Tools** sẽ xuất hiện mục **Data Analysis**.

➤ Quy trình để báo số để thực hiện cài thêm **Moving Average**

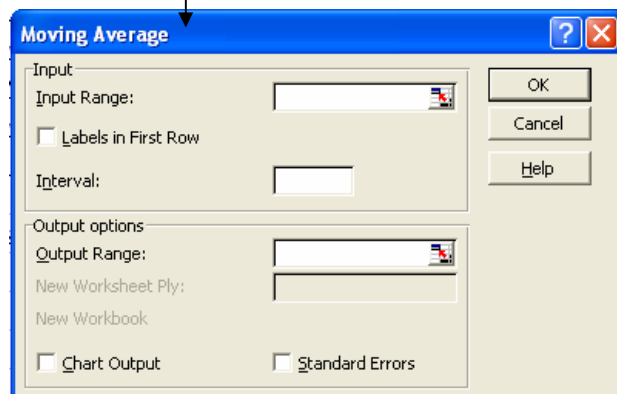
+ Nhập số liệu thu thập được vào bảng tính

+ Chọn **Tools | Data Analysis | Moving Average, OK**. Các hộp thoại liên tiếp xuất hiện như hình sau:



Hình 4.4 Hộp thoại chọn các công cụ phân tích dữ liệu

Hình 4.5 Hộp thoại **Moving Average**



### **M t s thu t ng :**

**Input Range:** Vùng a ch ch a các quan sát ã bi t

**Labels in First Row:** Tích vào ây kh ng nh ô u tiên c ch n không ch a d li u.

**Interval:** là n k tr c k d báo.

**Output Option:** Khai báo vùng k t xu t k t qu .

**Output Range:** Nh p vào vùng a ch ch a k t qu ho c a ch ô u tiên phía trên bên trái c a vùng ch a k t qu

**New Worksheet Ply:** K t qu c xu t ra trên m t sheet m i.

**New Workbook:** K t qu c xu t ra trên m t file Excel m i.

**Chart Output:** Tích vào m c này a ra th k t qu d báo.

**Standard Errors:** a ra các sai s chu n c a các d báo.

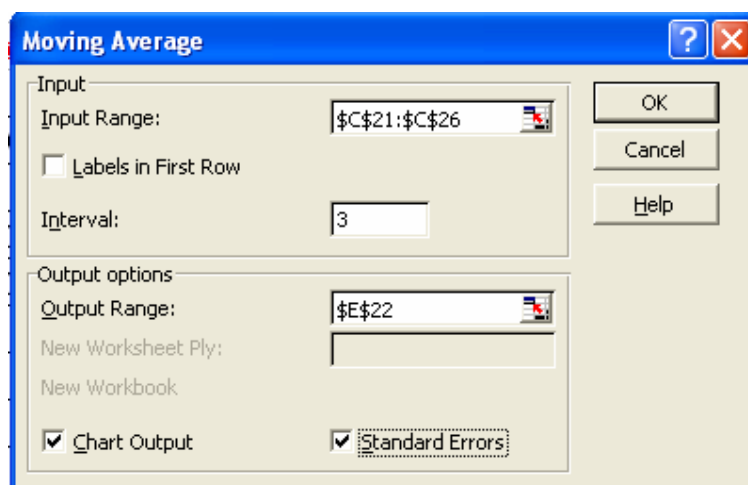
+ Nh n *OK* a ra k t qu d báo.

L i xét ví d 4.1 trên d báo b ng ph ng pháp s d ng trình cài thêm *Moving Average*. Các b c th c hi n nh sau:

- Nh p có s li u thu th p c vào b ng tính nh trên.

- *Tools \ Data Analysis \ Moving Average, OK*. B ng h p tho i xu t hi n ta

i n các thông tin vào nh hình sau:



Hình 4.6 Nh p các thông s cho mô hình d báo

- Nh n *OK* ta c b ng k t qu sau:

	A	B	C	D	E	F	G	H	I	J
17		<i>Phương pháp trung bình động dự báo số trẻ sơ sinh năm 2006 với n=3</i>								
18		<i>Sử dụng Moving Average</i>								
19										
20		<b>năm</b>	<b>số trẻ sơ sinh</b>	<b>dự báo</b>	<b>MA</b>	<b>sai số chuẩn</b>				
21		2000	29	-						
22		2001	30	-	#N/A	#N/A				
23		2002	28	-	#N/A	#N/A				
24		2003	31	29	29	#N/A				
25		2004	29	30	30	#N/A				
26		2005	27	29	29	0.981306763				
27		2006		<b>29</b>	<b>29</b>	1.401057801				
28										
29		<i>Nhận xét:</i>								
30		Dự báo đến năm 2006 số trẻ sơ sinh của địa phương A sẽ là 29 bé								
31		với sai số chuẩn là 1.4								
32		Đồ thị Moving Average cho ta hai đường:								
33		Đường thực tế Actual								
34		Đường dự báo Forecast								

#### 4.1.5 D báo b ng h i quy tuy n tính trong Excel

d báo h i quy tuy n tính trong Excel ta có r t nhi u cách nh s d ng các hàm c a Excel và s d ng trình cài thêm *Regression*.

##### 4.1.5.1 S d ng các hàm *TREND*, *FORECAST*, *LINEST*, *SLOPE* và *INTERCEPT*

d báo b ng ph ng pháp s d ng mô hình h i quy tuy n tính n y = ax + b (y là bi n ph thu c, x là bi n c l p) khi bi t c m t trong hai giá tr ta có th s d ng các hàm *TREND*, *FORECAST*, *LINEST*, *SLOPE* và *INTERCEPT*.

##### ➤ S d ng hàm *TREND*

- Tr v giá tr d c theo ng h i quy (theo ph ng pháp bình ph ng nh nh t)

- Cú pháp: =*TREND*(known\_y's, known\_x's, new\_x's, const)

- Trong ó: known\_y's, known\_x's, new\_x's là các giá tr ho c vùng a ch ch a giá tr ã bi t c a x, y t ng ng và giá tr m i c a x.

const là h ng s . Ng m nh n u const = 1 (True) thì h i quy theo hàm  $y = ax + b$ , n u const = 0 (False) thì h i quy theo hàm  $y = ax$ .



Xét ví dụ minh họa sau:

Ví dụ 4.2: Lợi nhuận của doanh nghiệp phụ thuộc vào giá thành sản phẩm. Dùng hàm *TREND* để báo lợi nhuận mà doanh nghiệp sẽ đạt được khi giá thành sản phẩm là 270.000 đồng. Ta có kết quả và công thức như sau:

	A	B	C	D	E	F
2		<i>Phương pháp dự báo hồi quy tuyến tính đơn <math>y=ax+b</math></i>				
3		<i>Sử dụng hàm <i>TREND</i></i>				
4						
5		<b>Lợi nhuận (tr đồng)</b>	<b>Giá thành</b>			
6		312	250000			
7		287	245000			
8		267	280000			
9		315	264000			
10		259	255000			
11		273	265000			
12		302	275000			
13		283	254000			
14		321	260000			
15		<b>288.811.356,9</b>	<b>270000</b>			
16						
17		B15=TREND(B6:B14,C6:C14,C15,1)				
18		<i>Nhận xét:</i>				
19		Lợi nhuận sẽ đạt 288.811.356,9 đồng khi giá thành là 270.000 đồng				

➤ **Sử dụng hàm *FORECAST***

- Tính, công việc giá trị tương lai dựa vào giá trị hiện tại.
- Cú pháp: =*FORECAST*(x, known\_y's, known\_x's)
- Trong đó: x là giá trị dùng để báo.

*known\_y's* là các giá trị hoặc vùng mà các dữ liệu phụ thuộc quan sát được

*known\_x's* là các giá trị hoặc vùng mà các dữ liệu phụ thuộc quan sát được.

Xét Ví dụ 4.2 trên: Ta có kết quả và công thức để báo lợi nhuận (y) tương ứng khi giá thành sản phẩm (x) là 270.000 đồng bằng hàm *FORECAST* như sau:

	G	H	I	J	K	L
2		<i>Phương pháp dự báo hồi quy tuyến tính đơn <math>y=ax+b</math></i>				
3		<i>Sử dụng hàm FORECAST</i>				
4						
5		<b>Lợi nhuận (tr đồng)</b>	<b>Giá thành</b>			
6		312	250000			
7		287	245000			
8		267	280000			
9		315	264000			
10		259	255000			
11		273	265000			
12		302	275000			
13		283	254000			
14		321	260000			
15		<b>288.8113569</b>	270000			
16						
17		H15=FORECAST(I15,H6:H14,I6:I14)				
18		<i>Nhận xét:</i>				
19		Lợi nhuận sẽ đạt 288.811.356,9 đồng khi giá thành là 270.000 đồng				

Như vậy, dù sử dụng hàm *TREND* hay hàm *FORECAST* đều cho ta các kết quả giống nhau.

### ➤ Sử dụng hàm *SLOPE* và *INTERCEPT*

Ngoài việc sử dụng hai hàm trên để báo cáo chúng ta cũng có thể sử dụng kết hợp hai hàm *SLOPE* tính hệ số góc  $a$  và hàm *INTERCEPT* tính hệ số  $b$  do biết các hàm hồi quy tuyến tính  $y=ax+b$ . Thay các hệ số  $a, b$  này vào hàm sẽ với giá trị  $x$  bất kỳ ta sẽ tìm ra giá trị còn lại cần báo.

- Cú pháp:  $=SLOPE(\text{known\_y's}, \text{known\_x's})$   
 $=INTERCEPT(\text{known\_y's}, \text{known\_x's})$

Trong đó: *known\_y's* là các giá trị hoặc vùng chứa các dữ liệu phải thu được quan sát được

*known\_x's* là các giá trị hoặc vùng chứa các dữ liệu cần lập quan sát được.

Xét Ví dụ 4.2 trên: Sử dụng hàm *SLOPE* và hàm *INTERCEPT* để báo mức lợi nhuận ( $y$ ) tối đa khi giá thành sản phẩm ( $x$ ) là 270.000 đồng như trong hình sau:

	N	O	P	Q	R	S
2		<i>Phương pháp dự báo hồi quy tuyến tính đơn <math>y=ax+b</math></i>				
3		<i>Sử dụng hàm SLOPE và hàm INTERCEPT</i>				
4						
5		<b>Lợi nhuận (tr đồng)</b>	<b>Giá thành</b>		<b>a</b>	<b>b</b>
6		312	250000		-0.000240217	353.6699277
7		287	245000			
8		267	280000		$y = -0.00024022 * x + 353.67$	
9		315	264000			
10		259	255000		y	<b>288.8113569</b>
11		273	265000			
12		302	275000		Công thức:	
13		283	254000		R6=SLOPE(O6:O14,P6:P14)	
14		321	260000		S6=INTERCEPT(O6:O14,P6:P14)	
15		?	270000		S10=R6*P15+S6	
16						
17		<i>Nhận xét:</i>				
18		Lợi nhuận sẽ đạt 288.811.356,9 đồng khi giá thành là 270.000 đồng				

### ➤ Sử dụng hàm LINEST

Ta có thể sử dụng hàm *LINEST* cho phương pháp dự báo mô hình hồi quy tuyến tính đơn  $y = ax + b$  và mô hình hồi quy tuyến tính bội  $y = a_1x_1 + a_2x_2 + \dots + a_nx_n + b$  (\*).

- Cú pháp: =*LINEST*((*known\_y's*, *known\_x's*, *const*, *stats*))

- Nhập xong công thức bằng tổ hợp phím *Ctrl + Shift + Enter*.

- Trong đó: *known\_y's*, *known\_x's*, là các giá trị hoặc vùng chứa các giá trị đã biết của x và y tương ứng.

*const* là hằng số. Nếu muốn nhập *const = 1* (True) thì tính toán hệ số *b*, nếu *const = 0* (False) bỏ qua *b* ( $b = 0$ ).

*stats* là các tham số thống kê. Nếu muốn nhập *stats = 1* thì tính các tham số thống kê, nếu *stats = 0* thì bỏ qua. Các tham số thống kê nhập

c tính bao gồm:

+ các hệ số của các biến độc lập theo thứ tự giảm dần  $m_n, m_{n-1}, \dots, m_2, m_1$ , *b* tức là  $a_n, a_{n-1}, \dots, a_2, a_1, b$  của mô hình (\*).

+ các sai số chuẩn của các hệ số  $se_n, se_{n-1}, \dots, se_2, se_1, se_b$  ( $se_b = \#N/A$  khi *const = False*).

+ hệ số xác định  $r^2$ , sai số của giá trị  $y$   $se_y$ .

+ phân phối  $F$ ,  $s_b$   $c$   $t$  do  $df$

+  $SS_{reg}$  (regression sum of square) và  $SS_{resid}$  (residual sum of

square). Bảng stats có bố trí như sau:

$a_n$	$a_{n-1}$	...	$a_2$	$a_1$	$b$
$se_n$	$se_{n-1}$	...	$se_2$	$se_1$	$se_b$
$r^2$	$se_y$				
$F$	$df$				
$SS_{reg}$	$SS_{resid}$				

- Thay các giá trị của các hệ số tìm được và các giá trị của các biến đã biết vào hàm hồi quy tìm ra giá trị cần dự báo.

Xét ví dụ 4.3: Lợi nhuận của doanh nghiệp ( $y$ ) phụ thuộc và giá thành sản phẩm ( $x_1$ ), chi phí quản lý ( $x_2$ ), chi phí bán hàng ( $x_3$ ). Dự báo lợi nhuận của doanh nghiệp tại  $x_1 = 600$ ,  $x_2 = 35$ ,  $x_3 = 25$  bằng hàm *LINEST* như hình sau:

	A	B	C	D	E	F
25		<i>Phương pháp dự báo hồi quy tuyến tính sử dụng hàm LINEST</i>				
26		Lợi nhuận y	Giá thành x1	Chi phí quản lý x2	Chi phí bán hàng x3	
27		(tr. đồng)	(nghìn đồng)	(tr. đồng)	(tr. đồng)	
28		540.000	450	30.2	20.3	
29		520.000	520	32.1	21.2	
30		541.000	550	28.7	22.1	
31		532.000	555	28.6	24.1	
32		530.000	525	31.2	26.3	
33		560.000	540	28.7	24.7	
34		547.000	560	29.3	23.5	
35		559.000	575	34.1	24.9	
36		592.000	580	34.5	24.6	
37		<b>733.364</b>	600	35.0	25.0	
38						
39		<i>LINEST</i>	<b>0.48183692</b>	<b>3.321121514</b>	<b>0.204214223</b>	<b>322.917329</b>
40			4.994794402	3.425372693	0.252185639	132.232015
41			0.371869528	21.45901741	#N/A	#N/A
42			0.986709887	5	#N/A	#N/A
43			1363.108415	2302.447141	#N/A	#N/A
44						
45		C39=LINEST(B28:B36,C28:E36,1,1)			Ctrl + Shift + Enter	
46		<i>Hàm hồi quy</i>				
47		$y = 0.481 * x_1 + 3.321 * x_2 + 0.204 * x_3 + 322.917$				
48		B37=C39*C37+D39*D37+E39*E37+F39				
49		<i>Nhận xét:</i> Khi giá thành là 600.000 đ, chi phí quản lý là 35.000.000 đ,				
50		chi phí bán hàng là 25.000.000 đ thì lợi nhuận doanh nghiệp sẽ đạt được là 733.364.000 đ				

*Chú ý: Trong trường hợp có hai biến ta cần tiến hành từng bước một trong trường hợp có nhiều biến trên.*

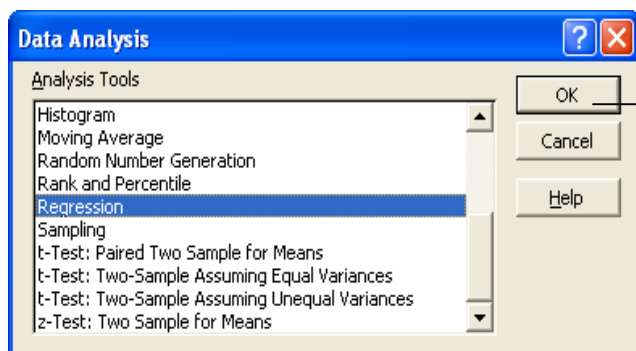
#### **4.1.5.2 Sắp đặt trình cài thêm Regression hồi quy và dự báo**

Ngoài việc sử dụng các hàm dự báo cho mô hình hồi quy tuyến tính như đã trình bày phần trên, ta có thể sử dụng trình cài thêm *Regression* trong bộ phân tích dữ liệu *Data Analysis*.

#### **Quy trình lắp ráp hồi quy tuyến tính trong Excel**

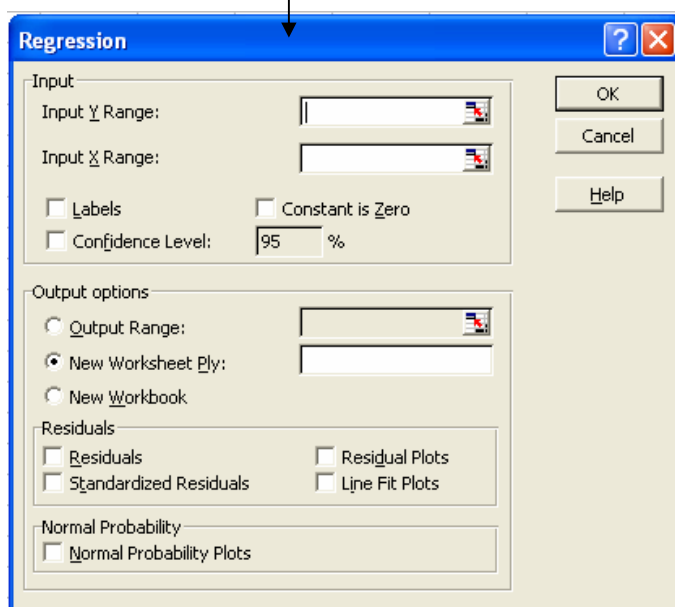
- Nhập số liệu vào bảng tính và ghi theo từng cột hoặc ghi theo từng dòng.

- Chọn *Tools \ Data Analysis \ Regression, OK*. Các bước tiếp theo như sau:



*Hình 4.7 Hình ảnh các công cụ phân tích dữ liệu*

*Hình 4.8 Hình ảnh khai báo các thông số của mô hình hồi quy*



### **M t s thu t ng :**

Các l a ch n nh p d li u vào *input*:

*Input Y Range*: Vùng a ch ch a bi n ph thu c Y

*Input X Range*: Vùng a ch ch a các bi n c l p X

*Labels*: Tích vào m c này kh ng nh ô (các ô) u tiên không ch a d li u h i quy

*Constant is Zero*: Tích vào m c này kh ng nh h s t do c a hàm h i quy tuy n tính  $b = 0$

*Confidence Level*: tin c y c a h i quy (m c nh là 95%) b ng  $1 - \alpha$  v i  $\alpha$  là m c ý ngh a hay xác su t m c sai l m lo i m t bác b  $H_0$  trong khi  $H_0$  úng.

Các l a ch n k t xu t k t qu *Output Option*:

*Output Range*: Vùng ho c ô phía trên bên trái c a vùng ch a k t qu

*New Worksheet Ply*: In k t qu ra m t sheet khác

*New Workbook*: In k t qu ra m t file Excel m i

Các l a ch n khác *Residuals*: Tích vào các m c này a ra

*Residuals*: Sai s do ng u nhiên

*Standardardized Residuals*: Chu n hoá sai s

*Residuals Plots*: th sai s

*Line Fit Plots*: th hàm h i quy tuy n tính

Xác su t phân ph i chu n *Normal Probability*:

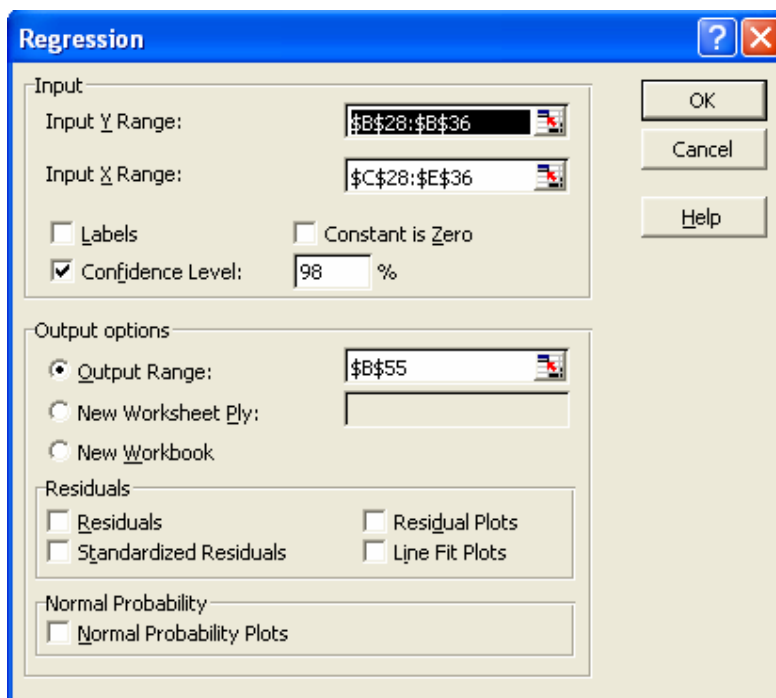
*Normal Probability Plots*: th xác su t phân ph i chu n

- Nh n *OK* a ra k t qu h i quy.

Thay các h s c a mô hình h i quy tính c và các giá tr ã cho trong k d báo vào hàm h i quy ta s tính c giá tr c n d báo.

*L i xét ví d 4.3 trên*: L i nhu n c a doanh nghi p (y) ph thu c và giá thành s n ph m ( $x_1$ ), chi phí qu n lý ( $x_2$ ), chi phí bán hàng ( $x_3$ ). D báo l i nhu n c a doanh nghi p t c khi  $x_1 = 600$ ,  $x_2 = 35$ ,  $x_3 = 25$  b ng công c *Regression* ta làm nh sau:

- Nhập số liệu vào bảng tính như phần trên
- Chọn *Tools\ Data Analysis\ Regression, OK*. Bảng hiển thị *Regression* xuất hiện ta điền các thông tin như trong hình sau:



Hình 4.9 Khai báo các thông số của mô hình

- Nhấn *OK* ta có bảng kết quả sau:

	B	C	D	E	F	G	H	I	J
53	<i>Phương pháp dự báo hồi quy sử dụng trình cài thêm REGRESSION</i>								
54									
55	<b>SUMMARY OUTPUT</b>								
56									
57	<i>Regression Statistics</i>								
58	Multiple R	0.6098111							
59	R Square	0.3718695							
60	Adjusted R Square	-0.005009							
61	Standard Error	21.459017							
62	Observations	9							
63									
64	<b>ANOVA</b>								
65		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
66	Regression	3	1363.108415	454.37	0.9867	0.46968689			
67	Residual	5	2302.447141	460.49					
68	Total	8	3665.555556						
69									
70		<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 98.0%</i>	<i>Upper 98.0%</i>
71	Intercept	322.91733	132.2320146	2.4421	0.0585	-16.99533	662.82999	-122.0342	767.86884
72	X Variable 1	0.2042142	0.252185639	0.8098	0.4549	-0.4440485	0.852477	-0.644373	1.0528013
73	X Variable 2	3.3211215	3.425372693	0.9696	0.3768	-5.4840649	12.126308	-8.205019	14.847262
74	X Variable 3	0.4818369	4.994794402	0.0965	0.9269	-12.35767	13.321344	-16.3253	17.288972

*M t s thu t ng trong b ng k t qu :*

+ B ng tóm t t *SUMMARY OUTPUT*:

*Regression Statistics*: Các thông s c a mô hình h i quy

*Multiple R*: H s t ng quan b i ( $0 \leq R \leq 1$ ). Cho th y m c ch t ch c a m i liên h t ng quan b i.

*R Square*: H s xác nh. Trong 100% s bi n ng c a bi n ph thu c Y thì có bao nhiêu % s bi n ng là do các bi n c l p X nh h ng, còn l i là do sai s ng u nhiên.

*Adjusted R*: H s xác nh m u i u ch nh. Là h s xác nh có tính n l n hay nh c a b c t do *df*.

*Standard Error*: Sai s chu n c a Y do h i quy.

*Observation*: S quan sát hay dung l ng m u.

+ B ng phân tích ph ng sai *ANOVA (Analysis of variance)*:

*Regression*: Do h i quy

*Residual*: Do ng u nhiên

*Total*: T ng c ng

*Df (Degree of freedom)*: S b c t do

*SS (Sum of Square)*: T ng bình ph ng c a m c ng (sai l ch) gi a các giá tr quan sát c a Y (ký hi u là  $Y_i$ ) và giá tr bình quân c a chúng

*MS (Mean of Square)*: Ph ng sai hay s bình quân c a t ng bình ph ng sai l ch k trên

*TSS ( Total Sum of Square)*: T ng bình ph ng c a t t c các m c sai l ch gi a các giá tr quan sát  $Y_i$  và giá tr bình quân c a chúng  $\bar{Y}$

Do h i quy *Regression ESS (Explained Sum of Square)* là t ng bình ph ng các sai l ch gi a các giá tr c a bi n ph thu c Y nh n c t hàm h i quy m u (ký hi u  $Y^*_i$ ). L n c a ESS ph n ánh m c giao ng c a các giá tr cá bi t c a mô hình v i giá tr trung bình m u hàm h i quy.



$$ESS = \sum_i (Y_i^* - Y)^2$$

Do ngẫu nhiên Residual *RSS (Residual Sum of Square)* là tổng bình phương của tất cả các sai lệch giữa các giá trị quan sát của  $Y (Y_i)$  và các giá trị nhúng của hàm hồi quy  $Y_i^*$

$$RSS = \sum_i e^2 = \sum_i (Y_i - Y_i^*)^2$$

Ta có thể kiểm tra chéo như sau:

$$TSS = ESS + RSS$$

$$R^2 = ESS / TSS$$

$$SD^2 = VAR = MSS \text{ of } RSS$$

*F-stat*: Tiêu chuẩn F dùng làm công cụ kiểm định tin cậy về mặt khoa học (thống kê) của toàn bộ phương trình hồi quy

*Significance F*: F lý thuyết

+ Bảng phân tích hồi quy:

*Coefficients*: Các giá trị của các hệ số hàm hồi quy:

- *Intercept*: Hệ số do b. Hệ số này cho thấy xu hướng phát hiện của hàm hồi quy

- *X Variable 1, X Variable 2, X Variable 3...* là các hệ số góc của các biến độc lập  $x_1, x_2, x_3...$

*Standard Error*: (se) là chuẩn của mẫu theo biến  $x_i$

*t-stat*: Tiêu chuẩn t dùng làm công cụ kiểm định tin cậy về mặt khoa học (thống kê) của các giả định  $a_i (i = 1, 2, 3, \dots, n)$  tức là các mối liên hệ giữa X và Y.

*P-value*: Xác suất  $t > t\text{-stat}$ , dùng kiểm định tin cậy về mặt khoa học (thống kê) của các giả định  $a_i (i = 1, 2, 3, \dots, n)$  tức là các mối liên hệ giữa X và Y.

*Lower 95%, Upper 95%, Lower 98%, Upper 98%*: là cận dưới và cận trên của khoảng chấp nhận cho các tham số với tin cậy 95% và tin cậy 98%.

- *Nhận xét*: Dựa vào bảng kết quả trên ta có phương trình hồi quy:

$$y = 0.204 * x_1 + 3.321 * x_2 + 0.482 * x_3 + 322.917$$

Như vậy khi  $x_1 = 600$ ,  $x_2 = 35$ ,  $x_3 = 25$  thì giá trị dự báo của  $y$  tính được là:  $y = 0.204 * 600 + 3.321 * 35 + 0.452 * 25 + 322.917 = 573.731$ . Tổng là 1 tỉ 573.731.000 đồng.

Ngoài ra, dựa vào bảng kết quả ta có thể thấy:

+ Nếu chi phí quản lý  $x_2$  và chi phí bán hàng  $x_3$  không đổi thì cứ tăng 1 nghìn đồng giá thành  $x_1$  sẽ làm cho lợi nhuận  $y$  tăng lên 0.204 triệu đồng.

+ Nếu giá thành  $x_1$  và chi phí bán hàng  $x_3$  không đổi thì cứ tăng 1 triệu đồng chi phí quản lý  $x_2$  sẽ làm cho lợi nhuận  $y$  tăng lên 3.321 triệu đồng.

+ Nếu giá thành  $x_1$  và chi phí quản lý  $x_2$  không đổi thì cứ tăng 1 triệu đồng chi phí bán hàng  $x_3$  sẽ làm cho lợi nhuận  $y$  tăng lên 0.482 triệu đồng.

+ Hệ số tự phát của mô hình  $b = 322.917$  cho thấy các nhân tố khác làm tăng lợi nhuận là 322.917 triệu đồng.

+ Multiple R = 0.61 cho thấy mối quan hệ giữa các biến là tương đối chặt chẽ.

+  $R^2 = 0.37$  cho thấy trong 100% số biến độc lập lợi nhuận thì có 37% biến động là do giá thành  $x_1$ , chi phí quản lý và chi phí bán hàng, còn 63% là do các yếu tố ngẫu nhiên và các yếu tố khác không có trong mô hình.

#### **4.1.5.3 Kiểm định các hệ số hồi quy và mô hình hồi quy**

Trong việc kiểm tra quan sát, sau khi xây dựng mô hình hồi quy ta sẽ tìm kiếm đúng của mô hình thống kê. Tuy nhiên, việc xác suất mắc sai lầm cho phép cần xác định mô hình liệu có phù hợp hay không? Khi nào thì cần điều chỉnh hệ số kiểm định sự phù hợp của mô hình, các hệ số hồi quy và kiểm tra các hệ số này về mặt tin cậy cho phép.

##### **➤ Kiểm định sự phù hợp của mô hình**

Có kết luận chính xác về sự phù hợp của mô hình (tốt hay không) mối quan hệ (sự tương quan) giữa các biến) ta tiến hành kiểm định trong các bước lý thuyết sau:

$H_0: R^2 = 0$  (không có mối quan hệ giữa các biến)

$H_1: R^2 \neq 0$  (có mối quan hệ giữa các biến)

Hoặc:

$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ ;

$H_1: \text{T ít nhất một hệ số } \beta_i \text{ khác } 0$ .

Tiêu chuẩn kiểm định dùng là:  $F = \frac{R^2 / (k-1)}{(1-R^2) / (n-k)}$

Nếu  $H_0$  đúng thì  $F$  có phân phối  $f_\alpha(k-1, n-k)$  cho nên miền bác bỏ vì vi phạm giả thiết  $H_0$  là:

$$W_\alpha = \left\{ F_{qs} = \frac{R^2 / (k-1)}{(1-R^2) / (n-k)}; F_{qs} > f_\alpha(k-1; n-k) \right\}$$

vì vi mô hình tuyến tính n ta luôn có  $R^2 = r^2_{XY}$ .

*Chú ý:*  $f_\alpha(k-1; n-k)$  có thể tính bằng cách sử dụng hàm *FINV*

Cú pháp: **=FINV(probability, deg\_freedom1, deg\_freedom2)**

Trong đó: *probability* là mức ý nghĩa, *deg\_freedom1*, *deg\_freedom2* là các số bậc tự do tương ứng (k-1) và (n-k) (vì k là số biến độc lập và n là số quan sát của mô hình hồi quy)

*Xét ví dụ 4.3:* Tiến hành kiểm định sự phù hợp của mô hình hồi quy mà ta làm như sau:

Kiểm định giả thuyết:

$H_0: R^2 = 0$  không có sự phụ thuộc giữa các biến

$H_1: R^2 \neq 0$  tồn tại ít nhất một hệ số phụ thuộc giữa các biến

Nhìn vào bảng kết quả ta thấy  $F_{qs} = 0.99$  và tính  $f_{0.02}(3, 6)$  theo công thức  $= \text{FINV}(0.02, 3, 6) = 7.29$  nên  $F_{qs} > f_{0.02}(3, 6)$  do đó  $F_{qs}$  thuộc miền bác bỏ  $W_{0.02}$ . Do đó bác bỏ  $H_0$  và chấp nhận  $H_1$  tức là có sự tương quan hay tồn tại mối quan hệ giữa các biến của mô hình (mô hình hồi quy này là phù hợp).

*Chú ý:* Ta cũng có thể kiểm định bằng cách so sánh giá trị F-crit (significance F) của mô hình với mức ý nghĩa  $\alpha$ .

➤ **Kiểm định giả thuyết về các hệ số hồi quy**

Sau khi tiến hành kiểm định phù hợp của mô hình ta cũng cần phải kiểm tra tổng hệ số trong mô hình hồi quy bằng cách kiểm định t để có thể nhận xét giá trị cụ thể nào đó của các hệ số.

Kiểm định về giả thuyết:  $H_0: \beta_i = \beta_0$  và  $H_1: \beta_i \neq \beta_0$

Tiêu chuẩn kiểm định: 
$$T = \frac{\hat{\beta}_i - \beta_0}{Se(\hat{\beta}_i)} \approx T(n-k).$$

Giá trị miền bác bỏ đối với  $H_0$  là: 
$$W_\alpha = \left\{ T_{qs} = \frac{\hat{\beta}_i}{Se(\hat{\beta}_i)}; |T_{qs}| > t_{\alpha/2}(n-k) \right\}$$

*Chú ý:*  $t_{\alpha/2}(n-k)$  có thể tính bằng cách sử dụng hàm *TINV*

Cú pháp: =*TINV(probability, deg\_freedom)*

Trong đó: *probability* là mức ý nghĩa ( $\alpha/2$ )

*deg\_freedom* là số bậc tự do ( $n - k$ )

*Xét ví dụ 4.3 trên:* Tiến hành kiểm định các hệ số hồi quy của mô hình hồi quy多元 với mức ý nghĩa  $\alpha = 0.05$  ta làm như sau:

Kiểm định về giả thuyết:  $H_0: a_1 = 0$  và  $H_1: a_1 \neq 0$

Tổng kết quả ta có:  $T_{qs} = 0.81$  và tính  $t_{0.025}(6)$  theo công thức =*TINV(0.025,6)* = 2.97 nên  $-t_{0.025}(6) < T_{qs} < t_{0.025}(6)$ . Do đó, chúng ta có thể bác bỏ  $H_0$  tức là mức ý nghĩa 5% chúng ta thấy bằng chứng rằng có sự khác biệt khác biệt thành phần như n

Tổng thể vậy vậy ta sẽ tiến hành kiểm định về ba hệ số còn lại của mô hình trong các trường hợp kiểm định trái và kiểm định phải.

➤ **Chọn lọc các hệ số hồi quy**

Khi đã xây dựng mô hình hồi quy多元 ta cần phải chọn lọc các hệ số của hàm hồi quy suy ra một cho từng trường.

Ta có thể nghĩ:  $T = \frac{\hat{\beta}_i - \beta_i}{Se(\hat{\beta}_i)} \sim T(n-k)$  với  $i=1, k$ .

Dựa vào thống kê này và các giá trị tới hạn Student ta có thể xây dựng khoảng tin cậy cho các hệ số quy tuyến tính sau:

Với tin cậy  $1-\alpha$  cho trước và giá trị tới hạn của quy luật phân phối Student ta luôn tìm được xác suất:

$$P\left(\left|\frac{\hat{\beta}_i - \beta_i}{Se(\hat{\beta}_i)}\right| < t_{\alpha/2}(n-k)\right) = 1-\alpha$$

biên giới tin cậy và rút ra được:

Khoảng tin cậy cho  $\beta$  tuyến tính là:

$$\left(\hat{\beta}_i - t_{\alpha/2}(n-k)Se(\hat{\beta}_i); \hat{\beta}_i + t_{\alpha/2}(n-k)Se(\hat{\beta}_i)\right)$$

*Chú ý:* Ta có thể dựa vào kết quả của Lower(1- $\alpha$ )% và Upper(1- $\alpha$ )% trên bảng kết quả biên giới khoảng tin cậy cho các hệ số quy.

*Xét ví dụ 4.3 trên:* Áp dụng công thức trên ta có bảng hệ số góc  $a_1$  cho mô hình tuyến tính với tin cậy  $1-\alpha = 98\%$  như sau:

$$t_{0.01}(6) = 3.707$$

$$0.204 - 3.707 \cdot 0.252 = -0.73 \text{ và } 0.204 + 3.707 \cdot 0.252 = 1.138$$

$$\text{nên } -0.73 < a_1 < 1.138$$

Vậy với tin cậy 98% hệ số quy giá thành và lợi nhuận trong khoảng vùng nghiên cứu giao nhau trong khoảng: (-0.73, 1.138)

Trong trường hợp trên ta lần lượt tiến hành kiểm tra cho các hệ số còn lại của mô hình hồi quy để suy ra kết quả cho từng biến.

#### 4.1.6 Dự báo bằng hồi quy phi tuyến

Các mô hình phi tuyến sau khi đã xác định dạng mô hình tuyến tính ta sẽ tiến hành hồi quy, kiểm tra và dự báo cho mô hình tuyến tính và nghiên cứu phân tích trên.

#### 4.1.6.1 Các mô hình phi tuyến có thể biến đổi về mô hình tuyến tính

biến đổi các mô hình phi tuyến về mô hình tuyến tính ta có thể sử dụng phương pháp logarit hai vế của phương trình, kết quả như sau. Sau đây là một số mô hình phi tuyến có thể biến đổi về mô hình tuyến tính:

##### ➤ Hàm sản xuất Cobb Douglas (CD)

$$Định hàm: Y = AX_1^{b_1} X_2^{b_2} \dots X_i^{b_i} \dots X_n^{b_n} \quad (1)$$

Trong đó: Y là kết quả sản xuất.  $X_1^{b_1}, X_2^{b_2}, \dots, X_i^{b_i}, \dots, X_n^{b_n}$  là mức đầu vào các yếu tố sản xuất (tài nguyên, công nghệ ...) cho sản xuất.

Đây là một hàm rất phù hợp với lý thuyết kinh tế về quy luật đầu tư thâm canh. Tính toán đơn giản vì có thể áp dụng tuyến tính bằng cách logarit hoá hai vế của (1):

$$\ln Y = \ln A + b_1 \ln X_1 + b_2 \ln X_2 + \dots + b_i \ln X_i + \dots + b_n \ln X_n$$

Ta có thể viết lại là:

$$\ln Y = b_0 + b_1 \ln X_1 + b_2 \ln X_2 + \dots + b_i \ln X_i + \dots + b_n \ln X_n \quad (2)$$

Đây chính là dạng mô hình tuyến tính với các biến là  $\ln Y$  (biến phụ thuộc),  $\ln X_1, \ln X_2, \dots, \ln X_i, \dots, \ln X_n$  (các biến độc lập).

Phân tích các tham số của hàm CD:

$$+ \text{Hệ số suất cận sản phẩm của yếu tố } i: \partial Y / \partial X_i = b_i \cdot \bar{Y} / \bar{X}_i \quad (i=1,2,\dots,n)$$

Có nghĩa là nếu đầu tư thêm 1 đơn vị của yếu tố sản xuất mang lại thêm  $\partial Y / \partial X_i$  đơn vị sản phẩm, với giả thiết là mức đầu vào các yếu tố khác không thay đổi.

+ Coefficient sản xuất theo yếu tố  $i: \eta_{Y X_i} = (\partial Y / \bar{Y}) / (\partial X_i / \bar{X}_i) = b_i \quad (i=1,2,\dots,n)$ . Có nghĩa là sản lượng tăng thêm  $b_i\%$  khi yếu tố sản xuất tăng thêm 1%, với giả thiết là mức đầu vào các yếu tố khác không thay đổi.

##### ➤ Hàm quy Parabol

Hàm quy Parabol là dạng phương trình bậc hai tam thức bậc 2:

$$Y = aX^2 + bX + c + U_i \quad \text{với } i = 1, 2, \dots, n$$

giải bài toán này sẽ có hai cách:

+ cách 1: dùng các tham số để dựng hàm hồi quy Parabol theo phương pháp bình phương nhỏ nhất:

$$f(a,b,c) = \sum_{i=1}^n (Y_i - aX_i^2 - bX_i - c)^2 \rightarrow \min$$

$$\text{Do đó: } \frac{\partial f}{\partial a} = 0; \quad \frac{\partial f}{\partial b} = 0; \quad \frac{\partial f}{\partial c} = 0;$$

$$\text{Hay: } a \sum X_i^4 + b \sum X_i^3 + c \sum X_i^2 = \sum X_i^2 Y_i$$

$$a \sum X_i^3 + b \sum X_i^2 + c \sum X_i = \sum X_i Y_i$$

$$a \sum X_i^2 + b \sum X_i + cn = \sum Y_i$$

Giải hệ phương trình ta xác định các hệ số của mô hình. Sau khi xác định xong các hệ số của mô hình ta sẽ viết được mô hình hồi quy.

+ với  $X_2 = X^2 = X \cdot X$  thì tiến hành cách 1 như vậy để mô hình hồi quy tuyến tính.

### ➤ Hồi quy Hyperbol

Hàm hồi quy Hyperbol sẽ có dạng:

$$Y = \frac{a}{X} + b + u_i \quad (i=1, 2, \dots, n).$$

giải bài toán này sẽ có hai cách:

+ cách 1: dùng các tham số để dựng hàm hồi quy Hyperbol theo phương pháp bình phương nhỏ nhất:

$$f(a,b) = \sum_{i=1}^n (Y_i - \frac{a}{X_i} - b)^2 \rightarrow \min \text{ Do đó: } \frac{\partial f}{\partial a} = 0; \quad \frac{\partial f}{\partial b} = 0;$$

$$\text{Hay: } a \sum \frac{1}{X_i^2} + b \sum \frac{1}{X_i} = \sum \frac{Y_i}{X_i} \text{ và } a \sum \frac{1}{X_i} + bn = \sum Y_i$$

Giải hệ phương trình ta tìm các hệ số a và b rồi thay trở lại phương trình hồi quy.

+ n giá trị cho vị trí cột trong Excel ta đặt  $Z = 1/X$  rồi tiến hành công việc tương tự như mô hình tuyến tính với hai biến  $Y$  và  $Z$ .

➤ **Hồi quy Hyperbol bội**

Hàm hồi quy Hyperbol bội có dạng:

$$Y = b_0 + \frac{b_1}{X_1} + \frac{b_2}{X_2} + \frac{b_3}{X_3} + \dots + \frac{b_n}{X_n}$$

chuyển về dạng hồi quy tuyến tính ta đặt  $Z_i = 1/X_i$  ta có phương trình công vi t là:

$$Y = b_0 + b_1 Z_1 + b_2 Z_2 + \dots + b_n Z_n.$$

Với mô hình tuyến tính này ta tiến hành các bước như mô hình tuyến tính nghiên cứu ph n trên.

➤ **Hồi quy mũ**

Hàm hồi quy mũ có dạng:  $Y = e^{b_0 + b_1 X_1 + b_2 X_2 + \dots + b_n X_n}$

Logarit cơ số e cho cả hai v ta có:

$$\ln Y = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_n X_n.$$

ây là mô hình hồi quy tuyến tính với biến phụ thuộc  $\ln Y$  và các biến công l p  $X_1, X_2, \dots, X_n$ .

➤ **Hồi quy dạng  $y = ba^x$**

Là dạng hàm mũ. Ta logarit cơ số e cho cả hai v ta có:

$$\ln Y = X \cdot \ln a + \ln b.$$

T s li u i u tra th c t ta tính các giá trị  $\ln$  s tr thành mô hình hồi quy tuyến tính n v i biến phụ thuộc  $\ln Y$  và biến công l p  $X$ .

*Xét ví d 4.4:* Ng i ta kh o sát và th m dò m i quan h c a n m i l i ng  $Y, X_1, X_2, X_3, X_4$  c bi t r ng m i ph thu c c a chúng có dạng phương trình sau:  $Y = b + a_1 * X_1 + a_2 * \ln X_2 + a_3 * X_3^2 + a_4 * 1/X_4$ . Với các s li u ã cho hãy h i quy mô hình và đ b áo  $Y$  khi  $X_1 = 20, X_2 = 15, X_3 = 50, X_4 = 8$  v i  $\alpha=0.05$

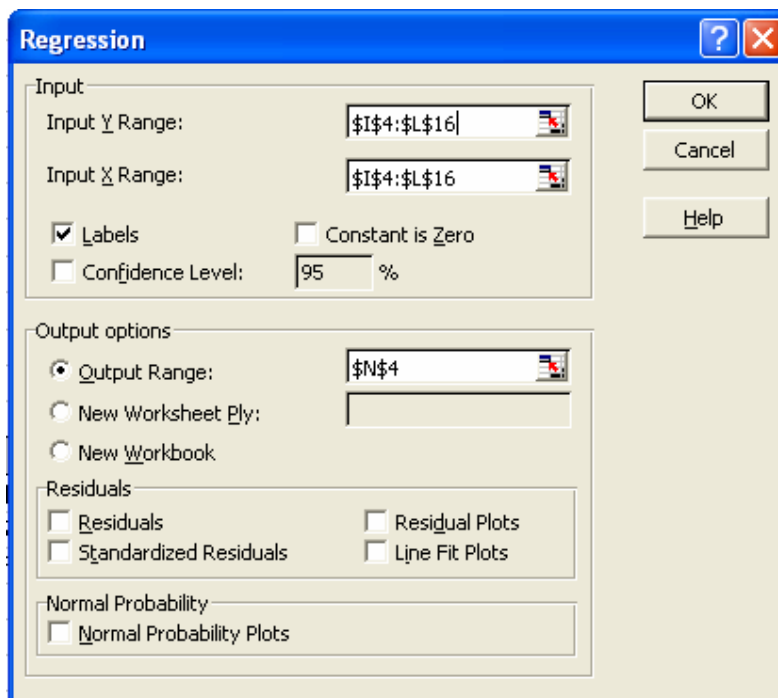
*Bài gi i:* Ta tiến hành theo các bước sau:

- Nh p, t và tính n ph cho các biến nh trong hình sau:



	B	C	D	E	F	G	H	I	J	K	L	
3	Hồi quy mô hình:			$Y=b+a1*X1 + a2*\ln X2+a3*X3^2+a4*1/X4$								
4	Y	X1	X2	X3	X4		Y	X1	LnX2	X3^2	1/X4	
5	613	15.6	12	28	4		613	15.6	2.4849	784	0.25	
6	50	18.5	6	46	3		50	18.5	1.7918	2116	0.3333	
7	13	16.2	12	35	5		13	16.2	2.4849	1225	0.2	
8	138	14.3	8	41	6		138	14.3	2.0794	1681	0.1667	
9	602	18.4	8	38	4		602	18.4	2.0794	1444	0.25	
10	914	15.6	17	48	5		914	15.6	2.8332	2304	0.2	
11	873	15.2	12	36	3		873	15.2	2.4849	1296	0.3333	
12	429	17.3	11	44	5		429	17.3	2.3979	1936	0.2	
13	102	16.8	10	26	5		102	16.8	2.3026	676	0.2	
14	115	16.4	12	50	6		115	16.4	2.4849	2500	0.1667	
15	134	14.3	8	46	4		134	14.3	2.0794	2116	0.25	
16	49	15.6	13	47	6		49	15.6	2.5649	2209	0.1667	
17		20	15	50	8			20	2.7081	2500	0.125	
18	Công thức:		H5=B5		J5=LN(D5)		copy công thức cho các ô còn lại					
19			I5=C5		K5=E5*E5							
20					L5=1/F5							

- Chọn *Tools\ Data Analysis\ Regression, OK*. Bảng hình phỏng diễn *Regression* xuất hiện ta điền các thông tin như trong hình sau:



Hình 4.10 Nhập các thông số cho mô hình

- Nhận OK ta có bảng kết quả sau:

	N	O	P	Q	R	S	T	U	V	
4	SUMMARY OUTPUT									
5										
6	<i>Regression Statistics</i>									
7	Multiple R	0.7307367								
8	R Square	0.5339761								
9	Adjusted R Square	0.2676767								
10	Standard Error	286.23894								
11	Observations	12								
12										
13	<i>ANOVA</i>									
14		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>				
15	Regression	4	657156.8918	2E+05	2.0052	0.19822				
16	Residual	7	573529.1082	81933						
17	Total	11	1230686							
18										
19		<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>	
20	Intercept	-2448.181	1620.720885	-1.51	0.1747	-6280.57	1384.212	-6280.574	1384.2125	
21	X1	3.2469211	66.96509075	0.048	0.9627	-155.1	161.5941	-155.1002	161.59409	
22	LnX2	836.04505	345.4971118	2.42	0.0461	19.07478	1653.015	19.07478	1653.0153	
23	X3^2	-0.017667	0.146049421	-0.12	0.9071	-0.36302	0.327685	-0.363019	0.3276846	
24	1/X4	3560.106	1664.83534	2.138	0.0698	-376.601	7496.813	-376.6012	7496.8133	
25										
26	<i>Nhận xét:</i>									
27	Phương trình hồi quy	Y = -2448.18 + 3.25*X1 + 836.05*LnX2 - 0.02*X3^2 + 3560.11*1/X4								
28	Với mức ý nghĩa 95% thì: khi X1 = 20, X2 = 15, X3 = 50, X4 = 8 giá trị dự báo Y là:	281.6551								
29	Công thức	U28=O20+I17*O21+J17*O22+K17*O23+L17*O24								
30										

#### 4.1.6.2 Sử dụng các hàm GROWTH và LOGEST

Ngoài việc sử dụng trình cài thêm Regression cho mô hình hồi quy phi tuyến ta còn có thể sử dụng hàm GROWTH và hàm LOGEST.

##### ➤ Sử dụng hàm GROWTH

- Dùng để hồi quy phi tuyến theo mô hình  $Y = b * m^X$

- Cú pháp: =GROWTH(known\_y's, known\_x's, new\_x's, const)

Trong đó: known\_y's, known\_x's, new\_x's là các giá trị hoặc vùng chứa giá trị đã biết của x, y tương ứng và giá trị mới của x.

const là hằng số. Nếu const = 1 (True) tính hằng số do b (nguyên), nếu const = 0 (False) bỏ qua hằng số b (b = 1).

Ví dụ 4.5: Giả sử giữa hai biến X và Y có mối quan hệ hàm mũ:  $Y = b * m^X$ . Ví dụ sau đây cho ta nhập vào bảng và tiến hành dự báo Y khi X = 20 như trong hình sau:

	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	<i>Dự báo phi tuyến bằng hàm GROWTH</i>												
2	Mô hình dự báo:			$Y = b * m^X$									
3													
4	Y	12.4	16.3	15.4	18.1	16.2	24.6	29.1	18.7	19.4	16.7	18.6	16.405
5	X	23	36	24	19	17	26	35	28	31	28	24	20
6													
7	Công thức:			R4=GROWTH(G4:Q4,G5:Q5,R5,1)									
8	Nhận xét:			Khi X = 20 thì Y = 16.405									

### ➤ S d ng hàm LOGEST

- Dùng h i quy phi tuy n theo mô hình:

$$Y = b * m_1^{X_1} * m_2^{X_2} * \dots * m_n^{X_n}$$

Cách s d ng hàm LOGEST gi ng nh hàm LINEST

- Cú pháp: =LOGEST(known\_y's, known\_x's, const, stat)

Trong ó: known\_y's, known\_x's, stat gi ng nh hàm LINEST.

*const* là h ng s . N u *const* = 1 (True) tính h s t do b (ng m nh), n u *const* = 0 (False) b qua h s b (*b* = 1).

N u b qua giá tr c a X thì gi thi t  $X = \{1, 2, 3, \dots\}$  v i s ph nt b ng s ph nt c a Y.

Ví d 4.6: Gi s gi a ba i l ng Y, X<sub>1</sub> và X<sub>2</sub> có m i quan h hàm m :  $Y = b * m_1^{X_1} * m_2^{X_2}$ . V i s li u ã cho ta nh p vào b ng tính và ti n hành d báo Y khi X<sub>1</sub> = 12 và X<sub>2</sub> = 25 nh trong hình sau:


	F	G	H	I	J	K	L	M	N	O	P	Q	R
14	<i>Dự báo phi tuyến bằng hàm LOGEST</i>												
15	Mô hình dự báo:			$Y = b * m_1^{X_1} * m_2^{X_2}$									
16	Y	X1	X2		LOGEST	0.97	1.02	412	Ctrl + Shift + Enter				
17	250	25	32			0.03	0.03	1.2					
18	243	16	28			0.18	0.42	#N/A					
19	261	13	35			0.86	8	#N/A					
20	189	14	34			0.3	1.38	#N/A					
21	452	23	29										
22	352	27	24		Công thức	L16=LOGEST(F17:F27,G17:H27,1,1)							
23	621	17	27		Hàm hồi quy:	Y = 412 * 1.02 ^X1 * 0.97 ^ X2							
24	531	21	23		Dự báo:	Y	khi	X1 = 12	X2 = 25				
25	184	19	26		Công thức	F28=N16*M16*G28*L16*H28							
26	253	16	21										
27	241	18	30		Như vậy, khi X1 = 12 và X2 = 25 thì Y = 123111.5								
28	123111.5	12	25										

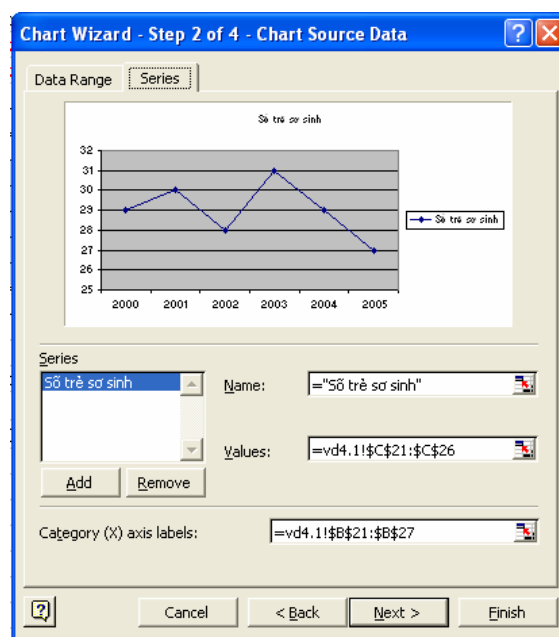
#### 4.1.7 Sử dụng tính năng báo

Ngoài việc sử dụng trình cài thêm *Moving Average* ta có thể báo trong Excel bằng cách tạo xu hướng như sau:

- + Chọn vùng số liệu thu thập (gọi là dòng cơ sở)
- + Khởi động *Chart Wizard* và làm theo các bước vẽ thể dạng *Line*
- + Sau khi vẽ xong, kích phẩy lên *series dữ liệu* và chọn *Add Strendline... \ Moving Average, OK* hoàn tất.

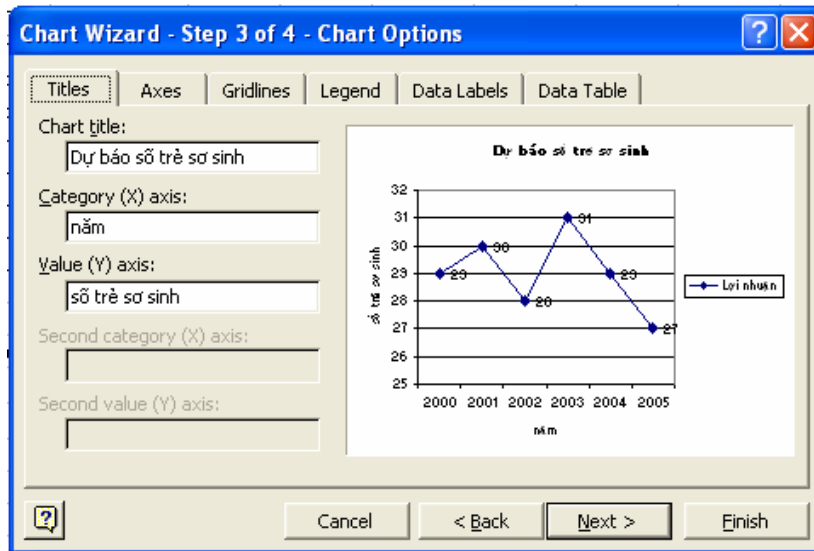
*Ví dụ 4.1 trên:* Vẽ thể báo số trẻ sơ sinh ảnh hưởng A theo các bước sau:

- + Chọn vùng số liệu thu thập C21:C26
- + Nhấp lên biểu tượng *Chart Wizard* trên thanh công cụ  hoặc chọn thành menu chính *Insert \ Chart...* Bảng hướng dẫn *Chart Wizard – step 1 of 4 – Chart Type* xuất hiện chọn loại thể *Line* trong mục *Standard Types* rồi nhấn *Next*.
- + Hướng dẫn *Chart Wizard – step 2 of 4 – Chart Source Data* (nguồn dữ liệu cơ bản) xuất hiện chọn *Columns* trong *Tab Data Range*. Trong *Tab Series* ta nhập các thông tin về tên thể *Name*, các giá trị (ảnh) của trục *X Category (X) axis labels* vào thể nh trong hình sau:



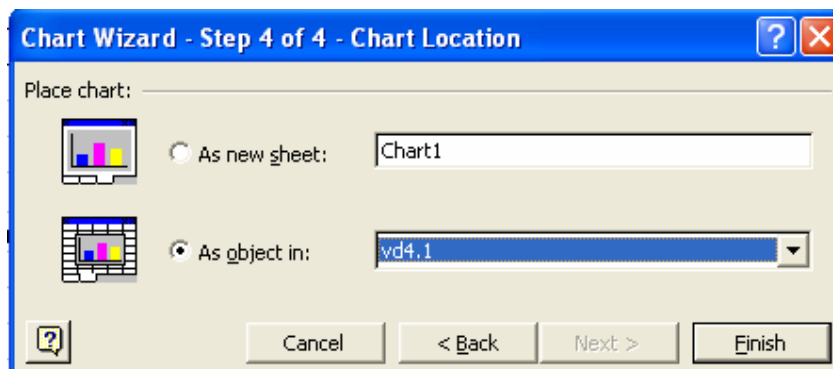
Hình 4.11 Hình ảnh thao tác khai báo nguồn dữ liệu

+ Nhấn *Next* chuyển sang hình ảnh thao tác tiếp theo *Chart Wizard – step 3 of 4 – Chart Options* gồm các lựa chọn cho thấy sau: *Titles* (tiêu đề), *Axes* (các trục), *Gridline* (ô lưới), *Legend* (vị trí các chú thích), *Data Labels* (các nhãn dữ liệu) và *Data Table* (bảng dữ liệu). Ta đặt tên cho các tiêu đề trong tab *Titles* như trong hình sau:



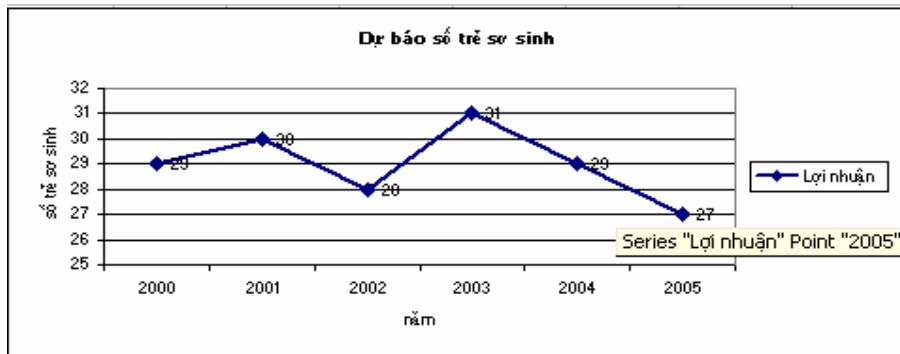
Hình 4.12 Hình ảnh thao tác chọn các lựa chọn

+ Nhấn *Next* chuyển sang hình ảnh thao tác lựa chọn vị trí thể hiện *Chart Wizard – step 4 of 4 – Chart Location*, tích vào *As new sheet* cho thể hiện thể vào một trang mới còn tích vào *As object in* thể hiện thể vào ứng dụng làm việc. Nhấn *Finish* vào *As object in* như hình sau:

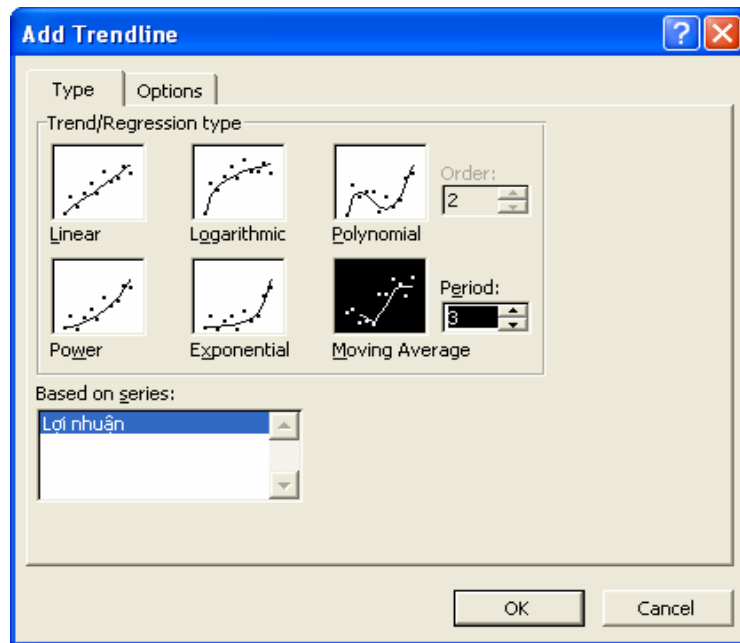


Hình 4.13 Hình ảnh thao tác xác định vị trí thể hiện

+ Nhấn *Finish* ta có thể nhìn hình sau:

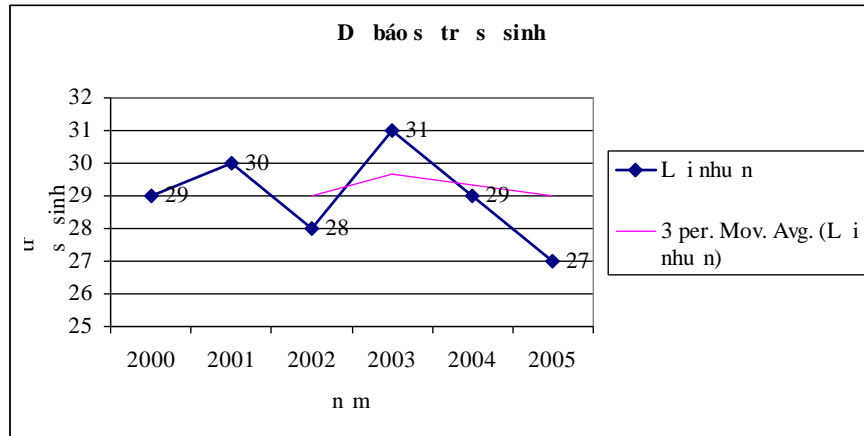


+ xu t hi n ng đ báo ta kích ph i vào ng th r i ch n Add Strendline, h p tho i Strendline xu t hi n ta ch n lo i th *Moving Average* v i 3 k trong tab *Type* nh hình sau:



Hình 4.14 H p tho i ch a các lo i th

+ Nh n OK ta c th sau:



## 4.2 PHÂN TÍCH TƯƠNG QUAN TRONG EXCEL

### 4.2.1 Những khái niệm cơ bản về tương quan

Khi nghiên cứu mối phụ thuộc, chúng ta xét theo trình tự thứ tự của mối phụ thuộc thì có 2 loại phụ thuộc sau đây:

**Phụ thuộc hàm** (mối liên hệ hàm số): Hai biến ngẫu nhiên  $X$  và  $Y$  có cùng giá trị là phụ thuộc hàm số nếu tồn tại  $f$  sao cho  $Y=f(X)$  tức là khi biết giá trị  $X$  thì theo một quy tắc nào đó có thể xác định giá trị tương ứng của  $Y$ . Đây là sự phụ thuộc hoàn toàn chắc chắn.

**Phụ thuộc thống kê** (mối liên hệ thống kê): Hai biến ngẫu nhiên  $X$  và  $Y$  có cùng giá trị là phụ thuộc thống kê nếu biết giá trị của  $X$  ta vẫn có thể xác định quy luật phân phối xác suất có điều kiện của  $Y$ :  $F(y/X=x) = P(Y>y/X=x)$ . Đây là sự phụ thuộc không hoàn toàn chắc chắn tức là khi biết giá trị của biến thì làm cho hình thức liên quan biến này nhưng nó không có hình thức hoàn toàn quy tắc như biến này.

Phân tích tương quan chủ yếu dựa trên các phân tích hai biến ngẫu nhiên là hệ thống tương quan (trường hợp hai biến ngẫu nhiên), hệ thống tương quan biến và hệ thống tương quan riêng phần (trường hợp có nhiều hơn hai biến ngẫu nhiên).

#### 4.2.1.1 Hệ thống tương quan

##### ➤ Định nghĩa

Hệ số tương quan của hai biến ngẫu nhiên liên tục quan hệ tuyến tính giữa hai biến mà không phân biệt biến nào là biến phụ thuộc biến nào là biến độc lập.

Giả sử  $X$  và  $Y$  là hai biến ngẫu nhiên có  $V(X) > 0$  và  $V(Y) > 0$ , thì hệ số tương quan của hai biến  $X$  và  $Y$  xác định như sau:

$$\rho_{XY} = \frac{E[(X - E(X))(Y - E(Y))]}{\sqrt{V(X)V(Y)}} \quad (5.1)$$

(\*) Hệ số  $\rho_{XY}$  tính bằng công thức trên chính là giá trị của 2 biến  $X$  và  $Y$  với công thức chúng chia cho trung bình nhân của 2 phương sai của 2 biến

➤ **Tính chất của hệ số tương quan**

(1).  $\rho_{XY} = \rho_{YX}$

(2).  $|\rho_{XY}| \leq 1$

(3).  $\rho_{XY} = 0$  nếu  $X$  và  $Y$  độc lập với nhau.

(4).  $\rho_{XY} = \pm 1$  khi và chỉ khi  $X$  và  $Y$  có mối quan hệ phụ thuộc tuyến tính.

(5).  $\rho_{XY} > 0$  nếu  $X$  và  $Y$  có mối tương quan thuận

$\rho_{XY} < 0$  nếu  $X$  và  $Y$  có mối tương quan nghịch

(6).  $|\rho_{XY}|$  càng gần với 1 thì mối phụ thuộc càng chặt chẽ hay mối quan hệ tuyến tính càng rõ

Như vậy ta có thể dùng  $\rho_{XY}$  để đo lường sự phụ thuộc tuyến tính của hai biến ngẫu nhiên.

➤ **Hệ số tương quan mẫu:**

Giả sử ta có  $\{(X_i, Y_i)\}$  với  $i=1, n$  là một mẫu ngẫu nhiên hai chiều,  $m_i$  là tần số xuất hiện của cặp  $(X, Y)$  xác định như sau:



X:  $X_1, X_2, \dots, X_k$

Y:  $Y_1, Y_2, \dots, Y_k$

$m_i$ :  $m_1, m_2, \dots, m_k$  trong ó  $\sum_{i=1}^k m_i = n$

H s t ng quan m u c xác nh theo công th c:

$$r = \frac{\left[ \sum_{i=1}^k (X_i - \bar{X})(Y_i - \bar{Y}) \right] / n}{\sqrt{\left[ \sum_{i=1}^k (X_i - \bar{X})^2 \right] / n} \sqrt{\left[ \sum_{i=1}^k (Y_i - \bar{Y})^2 \right] / n}} = \frac{\overline{XY} - \bar{X}\bar{Y}}{\sqrt{\bar{X}^2 - (\bar{X})^2} \sqrt{\bar{Y}^2 - (\bar{Y})^2}} \quad (5.2)$$

Trong ó:

$$\bar{X} = \left( \sum_{i=1}^k m_i X_i \right) / n$$

$$\bar{Y} = \left( \sum_{i=1}^k m_i Y_i \right) / n$$

$$\bar{X}^2 = \left( \sum_{i=1}^k m_i X_i^2 \right) / n$$

$$\bar{Y}^2 = \left( \sum_{i=1}^k m_i Y_i^2 \right) / n$$

$$\overline{XY} = \left( \sum_{i=1}^k m_i x_i y_i \right) / n$$

Cho nên:

$$r_{xy} = \frac{\sum_{i=1}^k m_i x_i y_i / n - \frac{\sum_{i=1}^k m_i x_i \sum_{i=1}^k m_i y_i}{n^2}}{\sqrt{\left( \sum_{i=1}^k m_i x_i^2 \right) / n - \left( \frac{\sum_{i=1}^k m_i x_i}{n} \right)^2} \sqrt{\left( \sum_{i=1}^k m_i y_i^2 \right) / n - \left( \frac{\sum_{i=1}^k m_i y_i}{n} \right)^2}} \quad (5.3)$$

Như vậy, có thể viết lại công thức tính  $r_{XY}$  trong trường hợp các tần số  $m_i = 1$  như sau:

$$r_{xy} = \frac{\sum_{i=1}^k x_i y_i / n - \frac{\sum_{i=1}^k x_i \sum_{i=1}^k y_i}{n^2}}{\sqrt{\left( \sum_{i=1}^k x_i^2 / n - \left( \frac{\sum_{i=1}^k x_i}{n} \right)^2 \right) \left( \sum_{i=1}^k y_i^2 / n - \left( \frac{\sum_{i=1}^k y_i}{n} \right)^2 \right)}} \quad (5.4)$$

**Những nhận xét về giá trị của r (\*)**

<b>Tần số r</b>	<b>Mức quan hệ các chỉ số</b>
r=0	X, Y hoàn toàn độc lập với nhau
r=1	X, Y có quan hệ tuyến tính với nhau
$0.0 <  r  < 0.3$	Mức tương quan yếu
$0.3 <  r  < 0.5$	Mức tương quan trung bình
$0.5 <  r  < 0.7$	Mức tương quan tương đối chặt
$0.7 <  r  < 0.9$	Mức tương quan chặt
$0.9 <  r  < 1$	Mức tương quan rất chặt

#### 4.2.1.2 Hệ số tương quan bội và hệ số tương quan riêng phần

➤ **Hệ số tương quan bội:** đánh giá mức độ chặt chẽ của mối liên hệ giữa một tiêu thức (thường là tiêu thức kết quả) với các tiêu thức còn lại (thường là tiêu thức nguyên nhân) có điều kiện loại trừ ảnh hưởng của các tiêu thức nguyên nhân khác

$$R_{y x_1 x_2 \dots x_n} = \sqrt{1 - \frac{\sum (y - \bar{y}_{x_1 x_2 \dots x_n})^2}{\sum (y - \bar{y})^2}} \quad (5.5)$$

Ví dụ 5.9: Trong ví dụ trình bày ở phần trước, giá trị thu được của 3 biến ngẫu nhiên  $Y$  là biến ngẫu nhiên và  $X_1, X_2$  là biến ngẫu nhiên độc lập có thể tính theo công thức sau:

$$R_{yx_1x_2} = \sqrt{1 - \frac{\sum (y - \bar{y}_{x_1x_2})^2}{\sum (y - \bar{y})^2}} \quad (5.6)$$

Hoặc có thể tính theo công thức sau:

$$R_{yx_1x_2} = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2}} \quad (5.7)$$

Trong đó,  $r_{yx_1}, r_{yx_2}, r_{x_1x_2}$  là các hệ số tương quan có thể tính theo công thức (5.3).

Nhận xét:

$$+ 0 \leq R \leq 1$$

+  $R = 0$  không có mối quan hệ tuyến tính, các biến độc lập với nhau.

+  $R$  càng gần 1 thì mối quan hệ càng chặt chẽ.

+  $R = 1$  thì có mối quan hệ hàm số.

➤ **Hệ số tương quan riêng phần:** dùng để đánh giá trình chặt chẽ của mối liên hệ giữa tiêu thức kết quả với từng tiêu thức nguyên nhân với điều kiện loại trừ ảnh hưởng của các tiêu thức nguyên nhân khác. Như trong trường hợp có mối liên hệ giữa  $Y$  với  $X_1$  và  $X_2$  có thể tính:

+ Hệ số tương quan riêng giữa  $Y$  và  $X_1$  (loại trừ ảnh hưởng của  $X_2$ )

$$r_{yx_1(x_2)} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{\sqrt{(1 - r_{yx_2}^2)(1 - r_{x_1x_2}^2)}} \quad (5.8)$$

+ Hệ số tương quan riêng giữa  $Y$  và  $X_2$  (loại trừ ảnh hưởng của  $X_1$ )

$$r_{yx_2(x_1)} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{\sqrt{(1 - r_{yx_1}^2)(1 - r_{x_1x_2}^2)}} \quad (5.9)$$

Việc tính toán theo công thức áp dụng công thức là rất vất vả và dễ mắc sai sót. Khi phức tạp như vậy chúng ta có thể thực hiện một số thao tác ngắn gọn

bảng vì các số dòng trình cài thêm *Correlation* trong *Data Analysis* tính hệ số tương quan trong Excel.

#### 4.2.2 Lựa chọn bảng tính hệ số tương quan trong Excel

Lựa chọn bảng tính hệ số tương quan trong Excel ta thực hiện theo hai bước sau:

Bước 1: Chọn các dữ liệu trên bảng tính

Bước 2: Tiến hành phân tích tương quan

##### 4.2.2.1 Chọn các dữ liệu trên bảng tính Excel phân tích tương quan

Số liệu quan sát của các biến phân loại có bố trí theo hàng theo cột hoặc theo hàng theo cột. Vì mỗi cách bố trí theo hàng và theo cột sẽ khiến ta có những cách tiến hành công việc khác nhau trong quá trình phân tích.

Trong trường hợp số liệu có các tổng riêng nhau ta có thể coi chúng đều có tổng là 1 ( $m_i = 1$ ) mà vẫn cho cùng một kết quả.

Trong trường hợp số liệu có các tổng là khác nhau ta nên áp dụng cách bố trí số liệu có cùng tổng hoặc có tổng đều là 1 tránh hiểu lầm và tiện cho quá trình phân tích vì Excel hiểu rằng mỗi hàng (cột, dòng) đều chứa số liệu của cùng một biến.

Ví dụ 4.7: Cách bố trí số liệu khi có các tổng khác nhau cụ thể mô tả trong hình sau:

	P	Q	R	S	T	U	V	W
3		Bố trí lại số liệu khi có mi khác nhau						
4								
5		<b>Xi</b>	<b>Yi</b>	<b>mi</b>		<b>Xi</b>	<b>Yi</b>	
6		13.2	12.5	2		13.2	12.5	
7		16.4	14.3	3		13.2	12.5	
8						16.4	14.3	
9						16.4	14.3	
10						16.4	14.3	
11								

Sau khi thực hiện các bước tính toán hành phân tích tương quan.

#### 4.2.2.2 Tiến hành phân tích tương quan

Có hai phương pháp lập bảng tính hệ số tương quan trong Excel: tính toán thủ công sử dụng các hàm trong Excel và sử dụng trình cài thêm *Correlation*

##### a. Phương pháp tính toán thủ công các hàm trong Excel.

➤ **Trình lập tính hệ số tương quan cho hai biến có các biến số  $m_i$  là khác nhau**

tính hệ số tương quan cho hai biến trong trình lập các biến số khác nhau ta sử dụng hàm *SUMPRODUCT* và hàm *SQRT*.

*Ví dụ 4.8:* Có số liệu về tuổi nghề (năm) và năng suất lao động (sản phẩm) của 100 công nhân tại một xí nghiệp công lập vì nhau. Hãy cho biết giữa tuổi nghề và năng suất lao động có mối phụ thuộc lẫn nhau hay không

*Bài giải:* Nhập số liệu quan sát vào bảng tính và sử dụng tính toán trình bày trong bảng sau:

A	B	C	D	E	F	G	H	I	J	K	L
1	Tuổi nghề	NSLĐ	Tần số	$m_i \cdot X_i$	$m_i \cdot Y_i$	$m_i \cdot X_i^2$	$m_i \cdot Y_i^2$	$m_i \cdot X_i \cdot Y_i$	$X_i^2$	$Y_i^2$	$X_i \cdot Y_i$
2	<b>Xi</b>	<b>Yi</b>	<b>mi</b>								
3	1	3	8	8	24	8	72	24	1	9	3
4	3	12	9	27	108	81	1296	324	9	144	36
5	9	21	9	81	189	729	3969	1701	81	441	189
6	10	24	8	80	192	800	4608	1920	100	576	240
7	7	12	12	84	144	588	1728	1008	49	144	84
8	12	27	8	96	216	1152	5832	2592	144	729	324
9	4	9	10	40	90	160	810	360	16	81	36
10	8	21	15	120	315	960	6615	2520	64	441	168
11	11	19	10	110	190	1210	3610	2090	121	361	209
12	5	16	11	55	176	275	2816	880	25	256	80
13	65	148	89	701	1468	5963	31356	13419			
14											
15	Từ số:	31.2832	CT:	I13/100-E13*F13/100^2							
16	Mẫu số:	32.07202	CT:	SQRT((G13/100-(E13/100)^2)*(H13/100-(F13/100)^2))							
17	r =	0.975405	CT:	C15/C16							
18											
19	Ta có thể dùng hàm SUMPRODUCT để tính các giá trị tổng										
20											
21	E13 =SUMPRODUCT(D3:D12,B3:B12)										
22	F13 =SUMPRODUCT(D3:D12,C3:C12)										
23	G13 =SUMPRODUCT(D3:D12,J3:J12)										
24	H13 =SUMPRODUCT(D3:D12,K3:K12)										
25	I13 =SUMPRODUCT(D3:D12,L3:L12)										

Hình 4.15 Tính hệ số tương quan sử dụng hàm SUMPRODUCT và hàm SQRT

Nhận xét:  $r = 0.88909252$  thể hiện mối phụ thuộc giữa tuổi nghề và năng suất lao động là mối phụ thuộc không kê (mối liên hệ tương quan) thuận và chặt. Tức là, khi tuổi nghề tăng thì năng suất lao động cũng tăng tương ứng.

➤ **Trình bày cách tính hệ số tương quan cho hai biến có các tần số  $m_i$  là biến ngẫu nhiên**

tính hệ số tương quan ta sử dụng các hàm SQRT, CORREL và RSQ.

+ Hàm SQRT: nghiên cứu trong chương 1.

+ Hàm CORREL: Tính hệ số tương quan giữa hai chuỗi số liệu.

Cú pháp: =CORREL(array1, array2)

Vì array1, array2 là các chuỗi số liệu.

+ Hàm *RSQ*: Tính bình phương hệ số tương quan giữa hai chuỗi số.

Cú pháp:  $=RSQ(array1, array2)$

Vì array1, array2 là các chuỗi số.

*Xét ví dụ 4.9:* Giả sử có tài liệu về tuổi nghề (n m) và năng suất lao động (s n ph m) của 10 công nhân tại một xí nghiệp. Hãy xác định xem giữa tuổi nghề của công nhân và năng suất lao động của họ có mối liên hệ gì với nhau không?

*Bài giải:* Để xác định tính chất mối quan hệ giữa tuổi nghề và năng suất lao động với trình độ học vấn  $m_i$  là bằng nhau ( $m_i = 1$ ) ta tiến hành tính hệ số tương quan giữa chúng sử dụng các hàm *SQRT*, *CORREL* và *RSQ* như sau:

- Nhập số liệu vào bảng tính
- Sử dụng công thức tính hệ số tương quan

Cách nhập số liệu và tính toán như trong bảng sau:

	A	B	C	D	E	F	G	H	I
2	Tên CN	Tuổi nghề Xi	NSLĐ Yi	$Xi+Yi$	$Xi^2$	$Yi^2$			
3	A	1	3	3	1	9			
4	B	3	12	36	9	144			
5	C	8	21	168	64	441			
6	D	10	24	240	100	576			
7	E	7	12	84	49	144			
8	G	4	9	36	16	81			
9	H	9	21	189	81	441			
10	I	5	16	80	25	256			
11	K	11	19	209	121	361			
12	L	12	27	324	144	729			
13	Tổng	70	164	1369	610	3182			
14	Tỷ số:		22.1	CT:	D13/10-B13*C13/10^2				
15	Mẫu số:		24.308023	CT:	SQRT((E13/10-(B13/10)^2)*(F13/10-(C13/10)^2))				
16	r:		0.9091648	CT:	C14/C15				
17	Hoặc								
18	r:		0.9091648	CT:	CORREL(B3:B12,C3:C12)				
19	r^2 =		0.8265807	CT:	RSQ(B3:B12,C3:C12)				

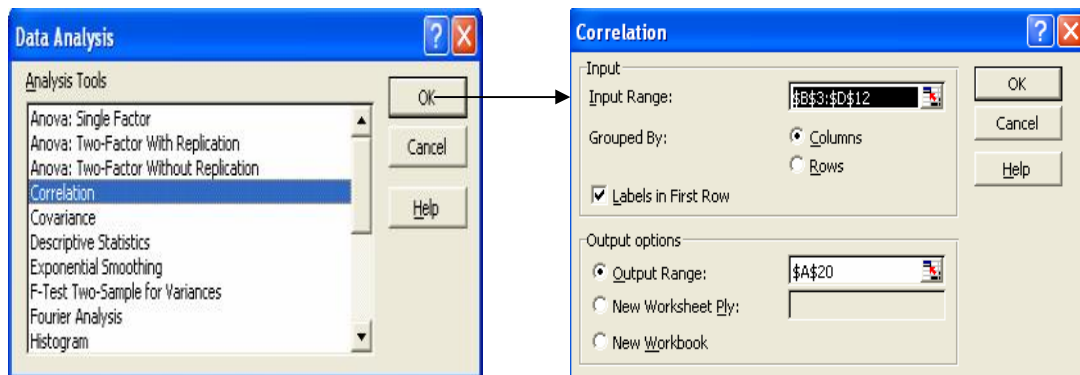
Hình 4.16 Tính hệ số tương quan sử dụng hàm *SQRT*, *CORREL* và *RSQ*

Nhận xét:  $r = 0.9091648$  nên có thể kết luận mối phụ thuộc giữa hai biến và ngược lại là mối phụ thuộc mạnh mẽ (mối liên hệ tương quan) và chính vì vậy nên có nghĩa là, khi biến này tăng lên thì biến kia cũng tăng lên tương ứng.

Như vậy, dù có tính toán hay không khi chúng ta tiến hành trình bày số liệu nhập dữ liệu không có tính toán thì vẫn có thể sử dụng hàm trong 2 hàm *RSQ* và *CORREL* tính hệ số tương quan cho tập số liệu.

### ***b. Phương pháp sử dụng trình cài thêm Correlation***

Chọn *Tools/ Data Analysis/ Correlation, OK*, hộp thoại xuất hiện và tiến hành thực hiện như sau:



Hình 4.17 Hộp thoại lựa chọn công cụ phân tích dữ liệu và hộp thoại khai báo các thông số lập bảng tính hệ số tương quan

### ***Một số chú ý***

+ *Input Range*: Khai báo vùng dữ liệu phân tích.

+ *Grouped by*: Chọn hình thức nghiên cứu theo:

+ *Columns*: X lý theo cột

+ *Rows*: X lý theo hàng

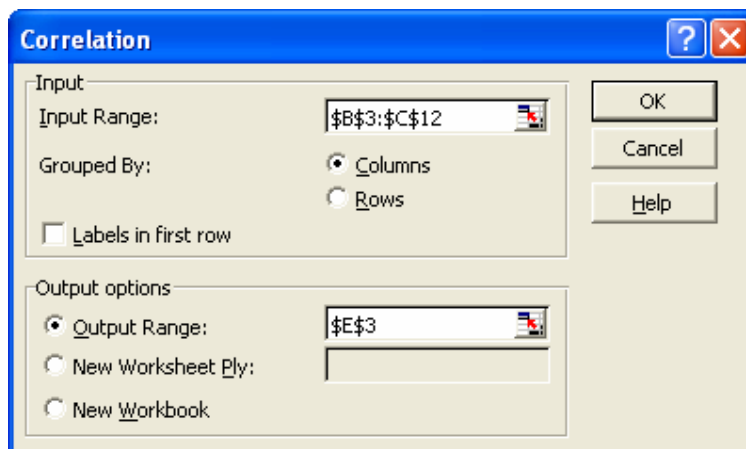
+ *Labels in First row*: Dòng đầu tiên trong vùng lấy làm tiêu đề.

+ *Output Range*: Khai báo vùng kết quả xuất ra.



Lưu ý ví dụ 4.9: Sau khi nhập dữ liệu vào bảng tính như mô tả trên ta tiến hành phân tích tương quan áp dụng trình cài thêm *Correlation* trong Excel như sau:

- Trên thanh công cụ hiển thị *Tools/ Data Analysis/ Correlation*, *OK*. Bảng thông tin *Correlation* xuất hiện ta nhập vào các thông tin như trong hình sau:



Hình 4.18 Khai báo các thông số của mô hình

- Nhấn *OK* ta có bảng kết quả sau:

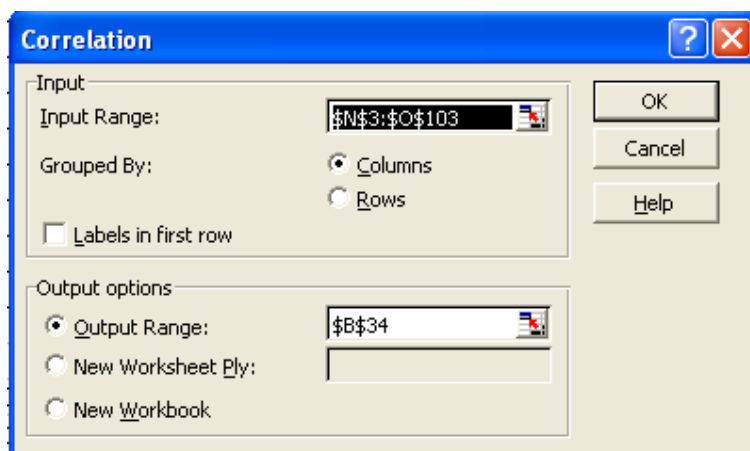
	A	B	C	D	E	F	G
2	Tên CN	Tuổi nghề Xi	NSLĐ Yi				
3	A	1	3			Xi	Yi
4	B	3	12		Xi	1	
5	C	8	21		Yi	<b>0.90916</b>	1
6	D	10	24				
7	E	7	12				
8	G	4	9				
9	H	9	21				
10	I	5	16				
11	K	11	19				
12	L	12	27				

- Nhận xét:  $r = 0.90916$  có cùng kết quả như khi tiến hành phân tích bằng phương pháp thống kê trên.

Xét ví dụ 4.8 trên: Sau khi bố trí số liệu về doanh số cùng tính là 1 (nằm trong vùng địa chỉ N3:O103) ta tiến hành phân tích tương quan bằng công cụ *Correlation* như sau:

- Tools/ Data Analysis/ Correlation, OK.

- Nhập vào thông tin bằng hình ảnh minh họa *Correlation* như trong hình sau:



Hình 4.19 Khai báo các thông số của mô hình

- Nhấn OK ta có bảng kết quả sau:

	A	B	C	D	E
33					
34			Column 1	Column 2	
35		Column 1	1		
36		Column 2	<b>0.8876539</b>	1	
37					

- Nhận xét: Dù làm theo phương pháp thủ công áp dụng các hàm hay theo phương pháp sử dụng *Correlation* thì đều cho cùng một kết quả  $r=0.8876539$ . Tức là giữa hai biến và ngược lại mối liên hệ tương quan thuận và chặt.

Ví dụ 4.10: Có tài liệu về doanh số lao động (nó tính là triệu đồng), phần trăm chi phí nguyên vật liệu chính trong giá thành sản phẩm và giá thành sản phẩm (nó tính là 1000 đồng) của xí nghiệp cùng sản xuất một loại sản phẩm như sau:

Thứ tự xí nghiệp	Giá thành đơn vị Y	NSL X <sub>1</sub>	%NVL chính X <sub>2</sub>
1	44	20	52
2	43	21	51
3	42	23	51
4	40	25	50
5	41	26	51

Bài gì i:

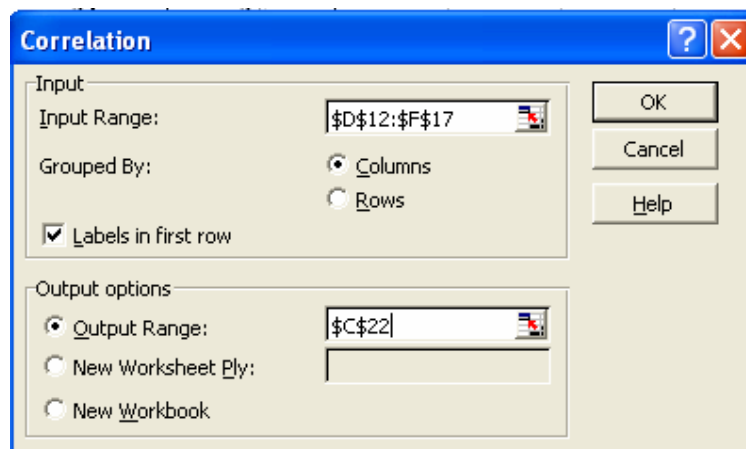
- Nhập số liệu vào bảng tính theo đúng cột C13:F17 như hình sau:

	B	C	D	E	F	G
10						
11		<b>Thứ tự xí nghiệp</b>	<b>Giá thành đơn vị Y</b>	<b>NSL X<sub>1</sub></b>	<b>%NVL chính X<sub>2</sub></b>	
12						
13		1	44	20	52	
14		2	43	21	51	
15		3	42	23	51	
16		4	40	25	50	
17		5	41	26	51	
18						

Hình 4.20 Nhập số liệu trên bảng tính

- Tools/ Data Analysis/ Correlation, OK.

- Nhập vào thông tin bảng tính để khai báo Correlation như trong hình sau:



Hình 4.21 Khai báo các thông số của mô hình

- Nhấn OK ta có bảng kết quả sau:

	B	C	D	E	F	G
21						
22			Y	X1	X2	
23		Y	1			
24		X1	-0.930260509	1		
25		X2	0.894427191	-0.693375245	1	
26						

Do đó:

$$R_{y_{x_1}, x_2} = \sqrt{\frac{(-0.94)^2 + 0.89^2 - 2(-0.94)0.89(-0.69)}{1 - (-0.69)^2}} = 0.985$$

-Nhấn xét:

Qua bảng phân tích tương quan ta thấy mối quan hệ giữa giá thành sản phẩm Y và năng suất lao động  $X_1$  và phần trăm nguyên vật liệu chính trong giá thành  $X_2$  là rất chặt ( $R_{y_{x_1}, x_2} = 0.985$ ) (\*) hoặc nhìn vào các hệ số tương quan  $r_{y_{x_1}} = -0.93026051$ ,  $r_{y_{x_2}} = 0.894427191$  và bảng (\*) trang 5. Trong mối quan hệ này thì năng suất lao động  $X_1$  tỉ lệ nghịch với giá thành sản phẩm Y, còn tỉ lệ giá trị nguyên vật liệu chính  $X_2$  tỉ lệ thuận với giá thành sản phẩm Y.

Bảng tính hệ số tương quan này cũng cho ta thấy có sự tương quan giữa các  $X_i$  hay không. Cần nhìn vào bảng kết quả ta thấy  $r_{x_1, x_2} = -0.6933752$  cho nên khi giá thành sản phẩm không phụ thuộc vào năng suất lao động  $X_1$  và phần trăm nguyên vật liệu chính trong giá thành  $X_2$  có mối quan hệ tương quan nghịch và chúng tỉ lệ nghịch với nhau.