



TS. NGUYỄN QUỐC HÙNG

- Mobile: 0912 251 253
- Email: hungngq@ueh.edu.vn
- Website: <https://bit.ueh.edu.vn/nqhung/>

KHOA HỌC DỮ LIỆU

- Mã học phần: **25D1INF50905948**
- Thời gian: **11/04/2025 - 16/05/2025**
- Hệ: ĐH, Chính quy
- Số lượng: 48 sinh viên
- Số tín chỉ: 2.00

PHƯƠNG PHÁP THÔNG KÊ, PHÂN TÍCH DỮ LIỆU

Thứ 6, thời gian: 12g45-17g05, Giảng đường: N1-303

NỘI DUNG

■ **Thống kê mô tả**

- Hàm, công cụ thống kê
- Báo cáo tổng hợp nhóm với Subtotal
- Hợp nhất dữ liệu với Consolidate
- Tổng hợp dữ liệu đa chiều với PivotTable

■ **Phân tích dự báo**

- Trung bình trượt (Moving Average)
- San bằng mũ (Exponential Smoothing)
- Hồi quy (Regression)

■ **Phân tích tối ưu**

- Lập mô hình tối ưu
- Sử dụng công cụ SOLVER để giải mô hình tối ưu

THỐNG KÊ MÔ TẢ

Hàm thống kê thông dụng

Tên hàm	Cú pháp và công dụng
MIN	MIN(danh-sách) : trả về giá trị nhỏ nhất của các số trong danh-sách Ví dụ: = MIN(6,4,2,8) → Kết quả: 2
MAX	MAX(danh-sách) : trả về giá trị lớn nhất của các số trong danh-sách Ví dụ: = MAX(6,4,2,8) → Kết quả: 8
SUM	SUM(x1 , x2 , ... , xn): trả về tổng các số x1 , x2 , ... , xn
SUMIF	SUMIF(vùng-chứa-điều-kiện, điều-kiện, [vùng-tính-tổng]): trả về tổng các ô trong vùng-tính-tổng tương ứng theo thứ tự với các ô trong vùng-chứa-điều-kiện thỏa điều-kiện. Nếu bỏ qua tham số vùng-tính-tổng nghĩa là vùng-tính-tổng = vùng-chứa-điều-kiện.
SUMIFS	SUMIFS(vùng-tính-tổng, vùng-chứa-điều-kiện-1,điều-kiện-1, [vùng-chứa-điều-kiện-2,điều-kiện-2],...) : trả về tổng các ô trong vùng-tính-tổng ứng với thứ tự các ô trong vùng-chứa-điều-kiện-1 thỏa điều-kiện-1 và các ô trong vùng-chứa-điều-kiện-2 thỏa điều-kiện-2,...

Hàm thống kê thông dụng

Tên hàm	Cú pháp và công dụng
AVERAGE	AVERAGE(danh-sách) : trả về giá trị trung bình của các số trong danh-sách. Ví dụ: = AVERAGE (6,4,2,8) → Kết quả: 5.75
AVERAGEIF	AVERAGEIF(vùng-chứa-điều-kiện, điều-kiện, [vùng-tính-trung-bình]) : trả về giá trị trung bình của các ô trong vùng-tính-trung-bình tương ứng theo thứ tự với các ô trong vùng-chứa-điều-kiện thỏa điều-kiện. Nếu bỏ qua tham số vùng-tính-trung-bình nghĩa là vùng-tính-trung-bình = vùng-chứa-điều-kiện
AVERAGEIFS	AVERAGEIFS(vùng-tính-trung-bình,vùng-chứa-điều-kiện-1,điều-kiện-1, [vùng-chứa-điều-kiện-2,điều-kiện-2],...) : trả về trung bình của các ô trong vùng-tính-trung-bình ứng với thứ tự các ô trong vùng-chứa-điều-kiện-1 thỏa điều-kiện-1 và các ô trong vùng-chứa-điều-kiện-2 thỏa điều-kiện-2,...

Hàm thống kê chuyên dụng

Tên hàm	Cú pháp và công dụng
COUNT	COUNT(danh-sách) : trả về số lượng các ô có kiểu số trong danh-sách Ví dụ: =COUNT(2,4,7, "A") → Kết quả: 3
COUNTA	COUNTA(danh-sách) : trả về số lượng các ô có nội dung khác rỗng trong danh-sách. Ví dụ: =COUNTA(2,4,7, "A") → Kết quả: 4
COUNTBLANK	COUNTBLANK(danh-sách) : trả về số lượng các ô có nội dung rỗng trong danh-sách.
COUNTIF	COUNTIF(vùng-chứa-điều-kiện, điều-kiện) : trả về số lượng các ô trong vùng-chứa-điều-kiện thỏa điều-kiện.
COUNTIFS	COUNTIFS(vùng-chứa-điều-kiện-1, điều-kiện-1, [vùng-chứa-điều-kiện-2, điều-kiện-2],...) : trả về số lượng các ô trong vùng-chứa-điều-kiện-1 thỏa điều-kiện-1 và các ô trong vùng-chứa-điều-kiện-2 thỏa điều-kiện-2,...
RANK	RANK(ô, vùng, kiểu) : trả về thứ tự xếp hạng của ô trong vùng; kiểu = 1: xếp hạng tăng dần, kiểu = 0 hoặc lờ đi: xếp hạng giảm dần

Hàm thống kê chuyên dụng

Tên hàm	Cú pháp và công dụng
Số trung vị	<p>Số trung vị của một dãy số là giá trị ở vị trí chính giữa khi dãy số đó được sắp xếp theo thứ tự tăng dần (hoặc ngược lại).</p> <ul style="list-style-type: none">Nếu số phần tử N của dãy số là lẻ thì số trung vị là số ở vị trí thứ $\frac{n+1}{2}$Nếu số phần tử N của dãy số là chẵn thì số trung vị là giá trị trung bình của 2 số ở vị trí thứ $\frac{n}{2}$ và $\frac{n}{2} + 1$ <p>Cú pháp: MEDIAN(Số-thứ-1, Số-thứ-2,...)</p> <p>Trong đó:</p> <ul style="list-style-type: none">Số-thứ-1: bắt buộc.Số-thứ-2,...: tùy chọn. <p>Ví dụ: = MEDIAN(1,3,4,8,2) → Kết quả là 3 = MEDIAN(1,3,4,8,2,5) → Kết quả là 3.5</p>

Hàm thống kê chuyên dụng

Tên hàm	Cú pháp và công dụng
Số yếu vị	<p>Số yếu vị là số có tần số xuất hiện nhiều nhất trong một dãy số.</p> <p>Cú pháp: Có 3 cách dùng</p> <p>MODE(Số-thứ-1, Số-thứ-2,...)</p> <p>MODE.MULT(Số-thứ-1, Số-thứ-2,...)</p> <p>MODE.SNGL(Số-thứ-1, Số-thứ-2,...)</p> <p>Trong đó:</p> <ul style="list-style-type: none">• Số-thứ-1: bắt buộc.• Số-thứ-2,...: tùy chọn.• Nếu dãy số không có giá trị lặp lại thì hàm thông báo lỗi #N/A• Hàm MODE.SNGL tương tự hàm MODE.• Hàm MODE.MULT được dùng khi dãy số có nhiều số yếu vị. Kết thúc hàm bằng tổ hợp phím Ctrl + Shift + Enter <p>Ví dụ:</p> <p>= MODE(1,3,2,5,2,3) → Kết quả là 3 (dãy số có số 2 và số 3 xuất hiện 2 lần nhưng 3 là số xuất hiện trước)</p> <p>= MODE.MULT(1,3,2,5,2,3) → Kết quả là 3,2</p>

Hàm thống kê chuyên dụng

Tên hàm	Cú pháp và công dụng
Phương sai	<p><u>Tính phương sai của mẫu (sample)</u> Cú pháp: VAR(Số-thứ-1, Số-thứ-2,...) VAR.S(Số-thứ-1, Số-thứ-2,...)</p> <p>Trong đó:</p> <ul style="list-style-type: none">• Số-thứ-1: bắt buộc.• Số-thứ-2,...: tùy chọn.• Nếu các đối số có chứa giá trị logic (True/False) hoặc chuỗi thì dùng hàm VARA. Khi đó True = 1, False = 0 và giá trị chuỗi = 0. <p><u>Tính phương sai của tổng thể (population)</u> Cú pháp: VARP(Số-thứ-1, Số-thứ-2,...) VAR.P(Số-thứ-1, Số-thứ-2,...)</p> <p>Trong đó:</p> <ul style="list-style-type: none">• Số-thứ-1: bắt buộc.• Số-thứ-2,...: tùy chọn.• Nếu các đối số có chứa giá trị logic (True/False) hoặc chuỗi thì dùng hàm VARPA. Khi đó True = 1, False = 0 và giá trị chuỗi = 0.

Tính phương sai của mẫu, công thức:

Sample Variance (s^2)

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

s^2 = variance

x_i = term in data set

\bar{x} = Sample mean

\sum = Sum

n = Sample size

Minh họa các bước Tính phương sai

- Cho một mẫu dưới đây:

Sample data set

	X
X_1	17
X_2	15
X_3	23
X_4	7
X_5	9
X_6	13

Bước 1: Tính giá trị trung bình của mẫu

Sample mean (\bar{x})

$$\bar{x} = \frac{\sum x}{n}$$

$$\bar{x} = \frac{17 + 15 + 23 + 7 + 9 + 13}{6}$$

$$= \frac{84}{6}$$

$$= 14$$

	x	\bar{x}
x_1	17	
x_2	15	
x_3	23	
x_4	7	
x_5	9	
x_6	13	
Σ	84	14

Bước 2: Lấy từng điểm dữ liệu trừ đi giá trị trung bình

$$\begin{aligned} & \mathbf{x_i - \bar{x}} \\ x_1 - \bar{x} &= 17 - 14 = 3 \\ x_2 - \bar{x} &= 15 - 14 = 1 \\ x_3 - \bar{x} &= 23 - 14 = 9 \\ x_4 - \bar{x} &= 7 - 14 = -7 \\ x_5 - \bar{x} &= 9 - 14 = -5 \\ x_6 - \bar{x} &= 13 - 14 = -1 \end{aligned}$$

	X	\bar{X}	$x_i - \bar{x}$
x_1	17		3
x_2	15		1
x_3	23		9
x_4	7		-7
x_5	9		-5
x_6	13		-1
Σ	84	14	

Bước 3: Bình phương mọi kết quả

$(x_i - \bar{x})^2$		x	\bar{x}	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
$(x_1 - \bar{x})^2 = 3^2 = 9$	x_1	17		3	9
$(x_2 - \bar{x})^2 = 1^2 = 1$	x_2	15		1	1
$(x_3 - \bar{x})^2 = 9^2 = 81$	x_3	23		9	81
$(x_4 - \bar{x})^2 = -7^2 = 49$	x_4	7		-7	49
$(x_5 - \bar{x})^2 = -5^2 = 25$	x_5	9		-5	25
$(x_6 - \bar{x})^2 = -1^2 = 1$	x_6	13		-1	1
	Σ	84	14		

Bước 4: Tìm tổng của các giá trị bình phương

$$\sum (x_i - \bar{x})^2$$

$$= (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + (x_4 - \bar{x})^2 + (x_5 - \bar{x})^2 + (x_6 - \bar{x})^2$$

$$= 9 + 1 + 81 + 49 + 25 + 1$$

$$= 166$$

	x	\bar{x}	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
x_1	17		3	9
x_2	15		1	1
x_3	23		9	81
x_4	7		-7	49
x_5	9		-5	25
x_6	13		-1	1
Σ	84	14		166

Bước 5: Tính phương sai mẫu

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$$

$$n = 6$$

$$= \frac{166}{6-1}$$

$$= \frac{166}{5}$$

$$= 33.2$$

Hàm thống kê chuyên dụng

Tên hàm	Cú pháp và công dụng
Độ lệch chuẩn	<p>Độ lệch chuẩn là căn bậc 2 của phương sai.</p> <p><u>Tính độ lệch chuẩn của mẫu (sample)</u></p> <p>Cú pháp: STDEV(Số-thứ-1, Số-thứ-2,...) STDEV.S(Số-thứ-1, Số-thứ-2,...)</p> <p>Trong đó:</p> <ul style="list-style-type: none">• Số-thứ-1: bắt buộc.• Số-thứ-2,...: tùy chọn.• Nếu các đối số có chứa giá trị logic (True/False) hoặc chuỗi thì dùng hàm STDEVA. Khi đó True = 1, False = 0 và giá trị chuỗi = 0. <p><u>Tính độ lệch chuẩn của tổng thể (population)</u></p> <p>Cú pháp: STDEVP(Số-thứ-1, Số-thứ-2,...) STDEV.P(Số-thứ-1, Số-thứ-2,...)</p> <p>Trong đó:</p> <ul style="list-style-type: none">• Số-thứ-1: bắt buộc.• Số-thứ-2,...: tùy chọn.• Nếu các đối số có chứa giá trị logic (True/False) hoặc chuỗi thì dùng hàm STDEVPA. Khi đó True = 1, False = 0 và giá trị chuỗi = 0.

Thống kê bằng công cụ **Descriptive statistics**

► Cách thực hiện:

- ❑ Bước 1: Chuẩn bị bảng số liệu cần thống kê.
- ❑ Bước 2: Chọn lệnh **Data → Data Analysis → Descriptive Statistics**, xuất hiện hộp thoại **Descriptive Statistics**.
- ❑ Bước 3: Khai báo các thông số Input và lựa chọn các thông số **Output Options**.

Thống kê bằng công cụ Descriptive statistics

	A	B	C	D	E	F	G	H	I	J	K
1	Số liệu bán hàng trong tháng 03 tại siêu thị ABC như sau										
2											
3	Ngày	Gà (con)	Heo (kg)	Bò (kg)							
4	1	46	39	20							
5	2	55	41	17							
6	3	43	25	22							
7	4	48	30	31							
8	5	54	46	16							
9	6	65	36	40							
10	7	36	37	33							
11	8	40	23	18							
12	9	51	30	17							
13	10	53	33	26							
14	11	64	50	40							
15	12	32	44	28							
16	13	41	41	35							
17	14	46	28	30							
18	15	53	35	33							
19	16	47	37	27							

Descriptive Statistics

Input
Input Range:

Grouped By: ☒ Columns ☐ Rows

☒ Labels in first row

Output options
☒ Output Range:
☐ New Worksheet Ply:
☐ New Workbook

☒ Summary statistics
☒ Confidence Level for Mean: %
☐ Kth Largest:
☐ Kth Smallest:

OK Cancel Help

Ví dụ: Thống kê mô tả cho lượng thịt Heo (theo kg) bán được trong tháng 03 tại siêu thị ABC

Thống kê bằng công cụ **Descriptive statistics**

<i>Heo (kg)</i>	
Mean	35.9375
Standard Error	1.889595967
Median	36.5
Mode	41
Standard Deviation	7.558383866
Sample Variance	57.12916667
Kurtosis	-0.538643031
Skewness	0.03452706
Range	27
Minimum	23
Maximum	50
Sum	575
Count	16
Confidence Level(95	4.027578464

- Mean: Giá trị trung bình
- Standard Error: Sai số chuẩn
- Median: Trung vị
- Mode: Yếu vị
- Standard Deviation: Độ lệch chuẩn
- Sample Variance: Phương sai mẫu
- Kurtosis: Độ nhọn
- Skewness: Độ bất đối xứng (Độ nghiêng)
- Range: Khoảng biến thiên (Max – Min)
- Minimum: Số nhỏ nhất
- Maximum: Số lớn nhất
- Sum: Tổng
- Count: Số lượng phần tử
- Confidence Level: Độ tin cậy

Báo cáo tổng hợp nhóm với **Subtotal**

► Chức năng của **Subtotal**:

- ❑ Cho phép tổng hợp từng nhóm dữ liệu của các cột kiểu số trong cơ sở dữ liệu như: tìm tổng, số lớn nhất, số nhỏ nhất, số trung bình,...
- ❑ Kết quả tổng hợp được đặt trên hay dưới mỗi nhóm.

Báo cáo tổng hợp nhóm với Subtotal

OrderID	Salesperson	Order Date	Order Amount	Country
10248	Buchanan	16/07/2003	\$440.00	UK
10249	Suyama	10/07/2003	\$1,863.40	UK
10250	Peacock	12/07/2003	\$1,552.60	USA
10251	Leverling	15/07/2003	\$654.06	USA
10252	Peacock	11/07/2003	\$3,597.90	USA
10253	Leverling	16/07/2003	\$1,444.80	USA
10254	Buchanan	23/07/2003	\$556.62	UK
10255	Dodsworth	15/07/2003	\$2,490.50	UK
10256	Leverling	17/07/2003	\$517.80	USA
10257	Peacock	22/07/2003	\$1,119.90	USA
10258	Davolio	23/07/2003	\$1,614.88	USA
10259	Peacock	25/07/2003	\$100.80	USA
10260	Peacock	29/07/2003	\$1,504.65	USA
10261	Peacock	30/07/2003	\$448.00	USA
10262	Callahan	25/07/2003	\$584.00	USA
10263	Dodsworth	31/07/2003	\$1,873.80	UK
10264	Suyama	23/08/2003	\$695.62	UK
10265	Fuller	12/08/2003	\$1,176.00	USA
10266	Leverling	31/07/2003	\$346.56	USA

Cần thống kê
tổng số tiền
mà mỗi nhân
viên đã thực
hiện

Báo cáo tổng hợp nhóm với **Subtotal**

► Cách thực hiện:

- ❑ Bước 1: **Sắp xếp dữ liệu** theo cột muốn gom nhóm (Trong trường hợp này là Salesperson)
- ❑ Bước 2: Chọn toàn bộ cơ sở dữ liệu hay click chuột vào một ô bất kỳ trên dữ liệu.
- ❑ Bước 3: Chọn Data → Outline → Subtotal, xuất hiện hộp thoại Subtotal.

Báo cáo tổng hợp nhóm với Subtotal

Subtotal

At each change in:
Salesperson

Use function:
Sum

Add subtotal to:
☐ Country
☐ Salesperson
☒ Order Amount
☐ Order Date
☐ OrderID

☒ Replace current subtotals
☒ Page break between groups
☒ Summary below data

Remove All OK Cancel

- **At each change in:** chọn cột gom nhóm (Salesperson)
- **Use function:** chọn hàm thống kê dùng để tổng hợp dữ liệu (Sum)
- **Add subtotal to:** chọn cột thống kê giá trị (Order Amount)
- **Replace current subtotals:** chọn để thay thế kết quả thống kê trước đó
- **Page break between group:** chọn để ngắt trang giữa các nhóm, khi in ra giấy thì mỗi nhóm một trang giấy
- **Summary below data:** chọn để kết quả tổng hợp ở dưới mỗi nhóm
- **Remove All:** để xóa bỏ các kết quả tổng hợp

Báo cáo tổng hợp nhóm với Subtotal

1	2	3		A	B	C	D	E
	1			OrderID	Salesperson	Order Date	Order Amount	Country
+	44				Buchanan Total		\$68,792.25	
+	144				Callahan Total		\$123,032.67	
+	262				Davolio Total		\$182,500.09	
+	304				Dodsworth Total		\$75,048.04	
+	397				Fuller Total		\$162,503.78	
+	465				King Total		\$116,962.99	
+	591				Leverling Total		\$201,196.27	
+	743				Peacock Total		\$225,763.68	
+	809				Suyama Total		\$72,527.63	
-	810				Grand Total		\$1,228,327.40	

Hợp nhất dữ liệu với **Consolidate**

► Chức năng của **Consolidate**:

- ❑ Cho phép hợp nhất dữ liệu từ những bảng dữ liệu khác nhau.
- ❑ Consolidate có thể hợp nhất dữ liệu theo 2 hình thức:
 - Tổng hợp theo vị trí: các bảng dữ liệu giống nhau về cấu trúc.
 - Tổng hợp theo hạng mục (theo hàng và cột): các bảng dữ liệu khác nhau về cấu trúc.

Hợp nhất dữ liệu với Consolidate

	A	B	C	D	E	F	G
1	CÔNG TY ABC						
2	Cửa hàng số 1						
3	DOANH THU NĂM 2014						
4	Đơn vị: trăm triệu đồng						
5	Mặt hàng	Tháng 1	Tháng 2	Tháng 3	Tháng 4	Tháng 5	Tháng 6
6	Laptop	20	22	21	23	19	30
7	Máy tính bảng	11	12	14	13	15	17
8	Điện thoại	49	48	60	48	45	67

	A	B	C	D	E	F
1	CÔNG TY ABC					
2	Cửa hàng số 2					
3	DOANH THU NĂM 2014					
4	Đơn vị: trăm triệu đồng					
5	Mặt hàng	Tháng 1	Tháng 2	Tháng 3	Tháng 4	Tháng 5
6	Laptop	14	15	16	23	19
7	Máy tính bảng	15	17	18	17	20
8	Điện thoại	40	40	49	48	47
9	Máy ảnh	50	80	40	50	60

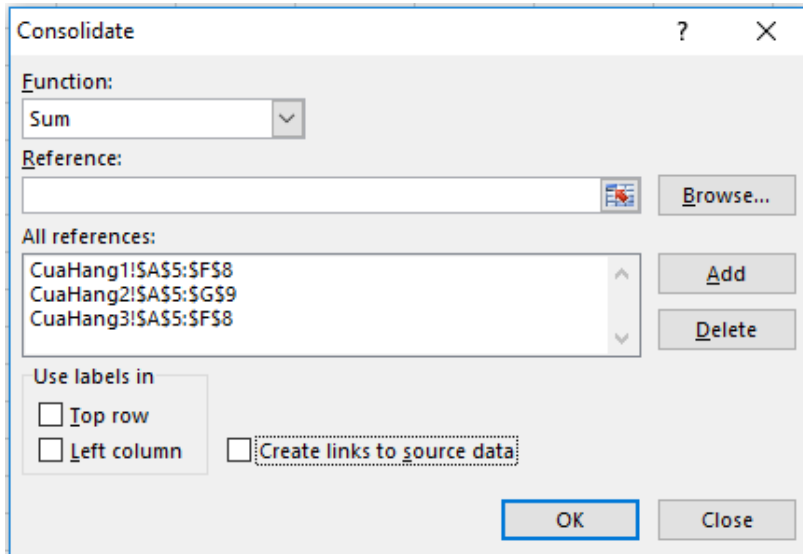
	A	B	C	D	E	F
1	CÔNG TY ABC					
2	Cửa hàng số 3					
3	DOANH THU NĂM 2014					
4	Đơn vị: trăm triệu đồng					
5	Mặt hàng	Tháng 1	Tháng 3	Tháng 4	Tháng 5	Tháng 6
6	Máy PC	20	22	23	19	20
7	Máy tính bảng	15	13	13	15	17
8	Máy ảnh	42	46	48	50	52

Hợp nhất dữ liệu với **Consolidate**

► Cách thực hiện:

- ❑ Bước 1: Chọn vùng sẽ chứa dữ liệu được hợp nhất.
- ❑ Bước 2: Chọn Data → Data Tools → Consolidate, xuất hiện hộp thoại Consolidate

Hợp nhất dữ liệu với Consolidate



The screenshot shows the 'Consolidate' dialog box in Microsoft Excel. It has a title bar with a question mark and a close button. The 'Function:' dropdown is set to 'Sum'. The 'Reference:' field is empty, with a 'Browse...' button next to it. The 'All references:' list contains three entries: 'CuaHang1!\$A\$5:\$F\$8', 'CuaHang2!\$A\$5:\$G\$9', and 'CuaHang3!\$A\$5:\$F\$8'. There are 'Add' and 'Delete' buttons next to this list. The 'Use labels in' section has three checkboxes: 'Top row' (unchecked), 'Left column' (unchecked), and 'Create links to source data' (checked). At the bottom are 'OK' and 'Close' buttons.

- **Function:** chọn hàm tổng hợp.
- **Reference:** để tham chiếu lần lượt các bảng dữ liệu nguồn.
- **All references:** chứa tất cả các vùng dữ liệu nguồn cần thiết cho việc hợp nhất.
- **Top row:** chọn nếu muốn dùng tên cột của vùng nguồn.
- **Left column:** chọn nếu muốn dùng các giá trị của cột đầu tiên của vùng nguồn.
- **Create links to source data:** chọn nếu muốn dữ liệu hợp nhất được cập nhật mỗi khi có thay đổi ở vùng dữ liệu nguồn.

Hợp nhất dữ liệu với Consolidate

→ Kết quả

	A	B	C	D	E	F	G
1		Tháng 1	Tháng 2	Tháng 3	Tháng 4	Tháng 5	Tháng 6
2	Laptop	34	37	37	46	38	30
3	Máy PC	20		22	23	19	20
4	Máy tính bảng	41	29	45	43	50	34
5	Điện thoại	89	88	109	96	92	67
6	Máy ảnh	92	80	86	98	110	52

Tổng hợp dữ liệu đa chiều với PivotTable

- ▶ Trong quá trình phân tích và tổng hợp dữ liệu, người ta thường có nhu cầu gom nhóm dữ liệu theo theo một số tiêu chí nào đó để dễ dàng quản lý.
- ▶ Excel cung cấp công cụ PivotTable để tổng hợp và phân tích dữ liệu với nhiều góc độ và nhiều cấp khác nhau.
- ▶ Nguồn dữ liệu để tạo PivotTable có thể là một cơ sở dữ liệu của Excel hay từ nguồn dữ liệu bên ngoài (như MS Access, MS SQL Server,...)

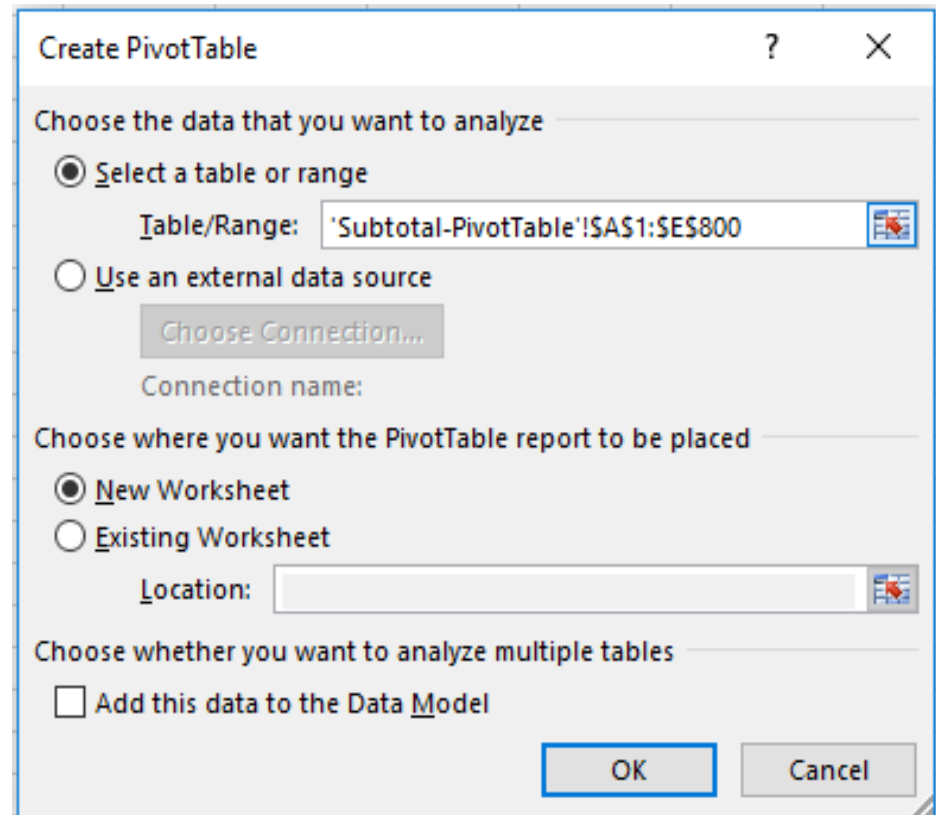
Tổng hợp dữ liệu đa chiều với PivotTable

Sum of Order Amount	Order Date			
Salesperson	2003	2004	2005	Grand Total
Buchanan	17.667	31.433	19.692	68.792
Peacock	49.945	124.656	51.163	225.764
Suyama	14.520	40.826	17.182	72.528
Grand Total	82.132	196.915	88.036	367.084

Tổng hợp dữ liệu đa chiều với PivotTable

► Cách thực hiện trên Excel:

- ❑ Bước 1: Click vào ô bất kỳ trên cơ sở dữ liệu.
- ❑ Bước 2: Chọn lệnh **Insert** → **PivotTable**.
- ❑ Bước 3: Xuất hiện hộp thoại **Create PivotTable**, chọn dữ liệu nguồn và nơi chứa **PivotTable**, click nút OK.



Tổng hợp dữ liệu đa chiều với PivotTable

► Cách thực hiện trên Excel:

- ❑ Bước 4: Drag các tên field từ **PivotTable Fields** vào 4 khu vực: **FILTERS**, **ROWS**, **COLUMNS** và **VALUES**.

The screenshot displays an Excel PivotTable and the PivotTable Fields task pane. The PivotTable is structured as follows:

Row Labels	UK	USA	Grand Total
Buchanan	68792.25		68792.25
Callahan		123032.67	123032.67
Davolio		182500.09	182500.09
Dodsworth	75048.04		75048.04
Fuller		162503.78	162503.78
King	116962.99		116962.99
Leverling		201196.27	201196.27
Peacock		225763.68	225763.68
Suyama	72527.63		72527.63
Grand Total	333330.91	894996.49	1228327.4

The PivotTable Fields task pane on the right shows the following configuration:

- Choose fields to add to report:** Country, Salesperson, Order Amount (checked); Order Date, OrderID (unchecked).
- Drag fields between areas below:**
 - FILTERS:** (Empty)
 - ROWS:** Salesperson
 - COLUMNS:** Country
 - VALUES:** Sum of Order Amount

Tổng hợp dữ liệu đa chiều với PivotTable

► Hiệu chỉnh PivotTable:

Thay đổi tên Field

- ❑ PivotTable tự tạo ra các ô có nền đậm khi thả các field vào vùng layout.
- ❑ Nội dung trong các ô này được tạo tự động dựa theo tên field và nơi bố trí.
- ❑ Có thể gỡ lại nội dung để báo cáo rõ ràng hơn.

Sum of Order Amou Column Lab			
Row Labels	UK	USA	Grand Total
Buchanan	68792.25		68792.25
Callahan		123032.67	123032.67
Davolio		182500.09	182500.09
Dodsworth	75048.04		75048.04
Fuller		162503.78	162503.78
King	116962.99		116962.99
Leverling		201196.27	201196.27
Peacock		225763.68	225763.68
Suyama	72527.63		72527.63
Grand Total	333330.91	894996.49	1228327.4

Tổng hợp dữ liệu đa chiều với PivotTable

► Hiệu chỉnh PivotTable:

Sắp xếp

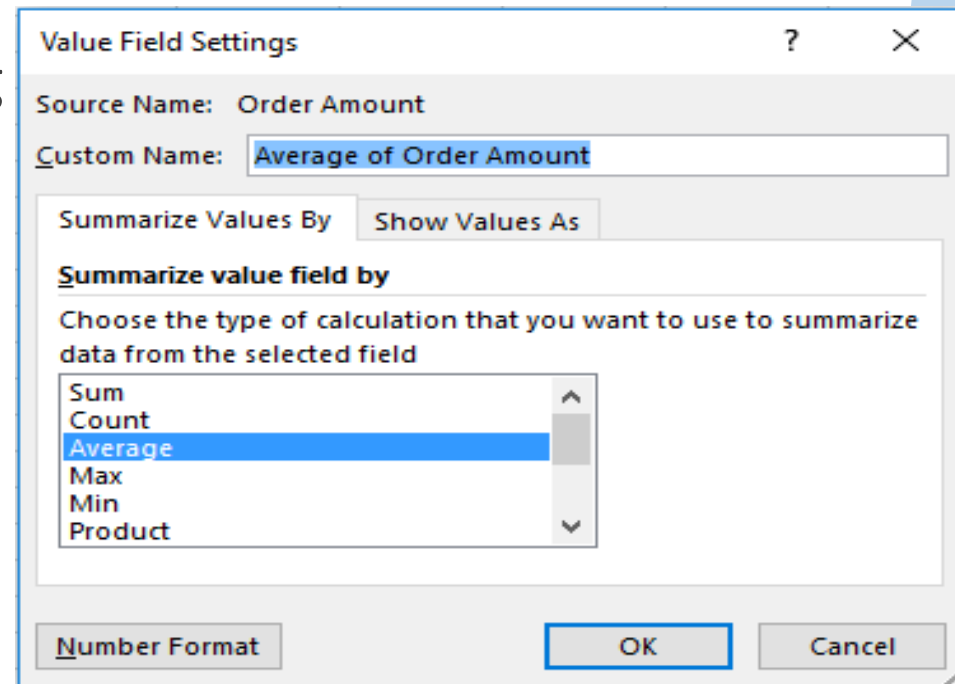
- ☐ Bấm vào mũi tên bên phải của nhãn dòng hoặc cột muốn sắp xếp.
- ☐ Bấm vào tùy chọn sắp xếp.

Sum of Order Amount	Column Labels	Row Labels	USA	Grand Total
23032.67		UK	82500.09	68792.25
82500.09		USA	75048.04	123032.67
62503.78		Grand Total	116962.99	182500.09
01196.27			225763.68	75048.04
25763.68			94996.49	162503.78
				116962.99
				201196.27
				225763.68
				72527.63
				1228327.4

Tổng hợp dữ liệu đa chiều với PivotTable

► Hiệu chỉnh PivotTable: Thay đổi phép tổng hợp dữ liệu

- ❑ Nhấn mũi tên tên xuống cạnh field cần định dạng trong khu vực **Values**
- ❑ Chọn **Value Field Settings** trong danh sách mở ra
- ❑ Chọn phương thức xử lý trong mục **Summarize value field by**



PHÂN TÍCH DỰ BÁO

Phương pháp: trung bình trượt (Moving Average)

► Công thức tính:

$$F_t = \frac{1}{w} \sum_{n=1}^{w \leq N} D_{t-w} = \frac{D_{t-1} + D_{t-2} + \dots + D_{t-w}}{w}$$

► Trong đó:

- F_t : là giá trị dự báo kỳ t
- D_t : là giá trị thực tế của kỳ t
- N: tổng số kỳ thực tế
- w: được gọi là cửa sổ trượt ($w \leq N$)

Phương pháp: trung bình trượt (Moving Average)

Cách thực hiện trên Excel:

- ▶ **Bước 1:** Chuẩn bị bảng số liệu cần dự báo
- ▶ **Bước 2:** Chọn lệnh **Data → Data Analysis → Moving Average**, xuất hiện hộp thoại **Moving Average**
- ▶ **Bước 3:** Khai báo các thông số **Input** và **Output Options**

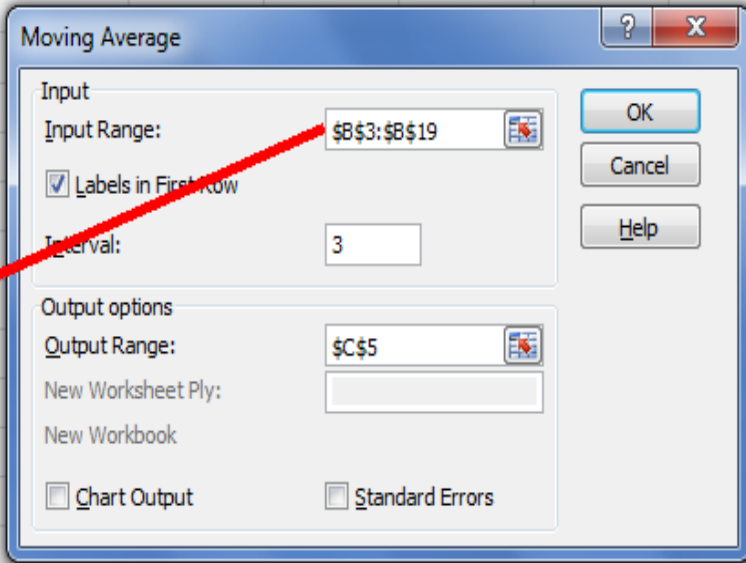
Phương pháp: trung bình trượt (Moving Average)

Cách thực hiện trên Excel:

- ❑ **Input Range** : tham chiếu đến vùng dữ liệu thực tế.
- ❑ **Labels in First Row**: Khai báo hàng đầu tiên của input range có chứa tiêu đề cột hay không.
- ❑ **Interval**: số lượng các kỳ trước đó muốn tính (w).
- ❑ **Output Range**: tham chiếu đến vùng xuất kết quả. Những ô không đủ số lượng các giá trị trước đó để tính toán sẽ nhận giá trị #N/A.
- ❑ **Chart Output**: tùy chọn dùng tạo biểu đồ nhúng cùng với vùng xuất kết quả.
- ❑ **Standard Errors**: tùy chọn dùng tạo thêm 1 cột chứa các sai số chuẩn.

Phương pháp: trung bình trượt (Moving Average)

	A	B	C	D	E	F	G	H	I
1	Số liệu bán hàng trong tháng 03 tại siêu thị ABC như sau								
2									
3	Ngày	Bò (kg)	Dự báo (W=3)						
4	1	20							
5	2	17	#N/A						
6	3	22	#N/A						
7	4	31	19.66666667						
8	5	16	23.33333333						
9	6	40	23						
10	7	33	29						
11	8	18	29.66666667						
12	9	17	30.33333333						
13	10	26	22.66666667						
14	11	40	20.33333333						
15	12	28	27.66666667						
16	13	35	31.33333333						
17	14	30	34.33333333						
18	15	33	31						
19	16	27	32.66666667						
20	17		30						



Phương pháp: San bằng mũ (Exponential Smoothing)

- ▶ **Công thức tính:** $F_t = a.D_{t-1} + (1-a).F_{t-1}$
- ▶ Trong đó:
 - F_t : là giá trị dự báo kỳ t
 - F_{t-1} : là giá trị dự báo kỳ t-1
 - D_{t-1} : là giá trị thực tế của kỳ t-1
 - a : là hệ số điều chỉnh (hệ số này tùy chọn sao cho $0 \leq a \leq 1$)

Phương pháp: San bằng mũ (Exponential Smoothing)

Cách thực hiện trên Excel:

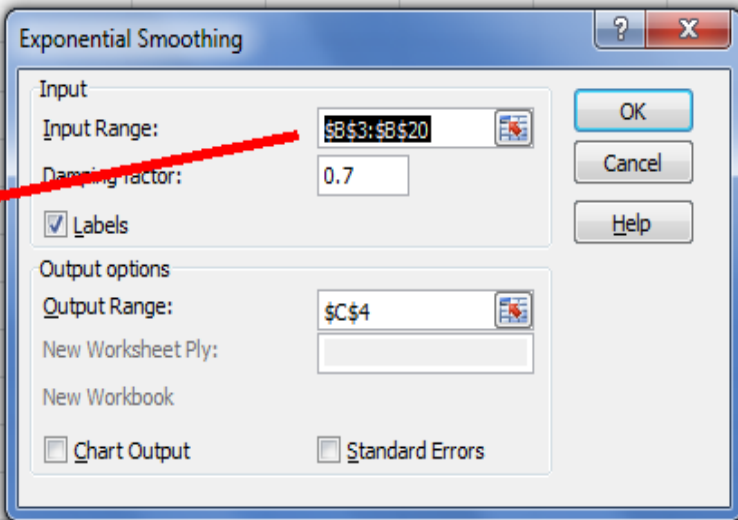
- ▶ **Bước 1:** Chuẩn bị bảng số liệu cần dự báo
- ▶ **Bước 2:** Chọn lệnh **Data → Data Analysis → Exponential Smoothing**, xuất hiện hộp thoại **Exponential Smoothing**
- ▶ **Bước 3:** Khai báo các thông số **Input** và **Output Options**

Phương pháp: San bằng mũ (Exponential Smoothing)

- ▶ **Input Range** : tham chiếu đến vùng dữ liệu thực tế.
- ▶ **Damping factor**: giá trị dùng làm hệ số san bằng. Đó là giá trị điều chỉnh sự bất ổn của dữ liệu, giá trị mặc định là **Damping factor $(1-a) = 0.7$**
- ▶ **Labels**: tùy chọn cho biết hàng/cột đầu tiên của input range có chứa tiêu đề hay không.
- ▶ Kết quả như trong hình sau là dự báo lượng thịt bò bán (kg) được tại siêu thị ABC vào ngày 17/03 với hệ số điều chỉnh $a = 0.3$ (Damping factor = 0.7). Như vậy lượng thịt bò dự báo cho ngày 17/03 là 29.88kg.

Phương pháp: San bằng mũ (Exponential Smoothing)

	A	B	C	D	E	F	G	H	I
1	Số liệu bán hàng trong tháng 03 tại siêu thị ABC như sau								
2									
3	Ngày	Bò (kg)	Dự báo ($\alpha = 0.3$)						
4	1	20	#N/A						
5	2	17	20.00						
6	3	22	19.10						
7	4	31	19.97						
8	5	16	23.28						
9	6	40	21.10						
10	7	33	26.77						
11	8	18	28.64						
12	9	17	25.45						
13	10	26	22.91						
14	11	40	23.84						
15	12	28	28.69						
16	13	35	28.48						
17	14	30	30.44						
18	15	33	30.31						
19	16	27	31.11						
20	17		29.88						



Phương pháp: Hồi quy (Regression)

- ▶ Phân tích hồi quy là nghiên cứu sự phụ thuộc của một biến (biến phụ thuộc hay còn gọi là biến được giải thích) vào một hay nhiều biến khác (biến độc lập hay còn gọi là biến giải thích)
- ▶ Phương trình hồi quy có dạng tổng quát:

$$Y = f(X_1, X_2, \dots, X_n)$$

(Hồi quy đơn biến: $Y = \mathbf{aX} + \mathbf{b}$)

- ▶ Trong đó:
 - Y: là biến phụ thuộc (dependent variable)
 - Xi: là các biến độc lập (independent variable)

Phương pháp: Hồi quy (Regression)

Cách thực hiện bằng hàm:

Tên hàm	Công dụng và cú pháp
TREND	<p>Trả về giá trị dọc theo đường hồi quy (theo phương pháp bình phương bé nhất)</p> <p>Cú pháp: TREND(known_y's, known_x's, new_x's, const)</p> <p>Trong đó:</p> <ul style="list-style-type: none">known_y's, known_x's, new_x's là các giá trị hoặc vùng địa chỉ chứa giá trị đã biết của x, y tương ứng và giá trị mới của x.Const là hằng số. Ngầm định nếu const = 1 (True) thì hồi quy theo hàm $y = a + bx$, nếu const = 0 (False) thì hồi quy theo hàm $y = bx$.

Phương pháp: Hồi quy (Regression)

Cách thực hiện bằng hàm:

Tên hàm	Công dụng và cú pháp
FORECAST	<p>Dự báo giá trị tương lai căn cứ vào các giá trị hiện tại. Cú pháp: FORECAST(x, known_y's, known_x's) Trong đó:</p> <ul style="list-style-type: none">• x là giá trị dùng để dự báo.• known_y's là các giá trị hoặc vùng địa chỉ của tập số liệu phụ thuộc quan sát được.• known_x's là các giá trị hoặc vùng địa chỉ của tập số liệu độc lập quan sát được

Phương pháp: Hồi quy (Regression)

Cách thực hiện bằng hàm:

Tên hàm	Công dụng và cú pháp
INTERCEPT	<p>Dùng để tính hệ số a của phương trình hồi quy đơn tuyến tính</p> <p>Cú pháp : INTERCEPT(known_y's, known_x's)</p> <p>Trong đó:</p> <ul style="list-style-type: none">known_y's là các giá trị hoặc vùng địa chỉ của tập số liệu phụ thuộc quan sát được.known_x's là các giá trị hoặc vùng địa chỉ của tập số liệu độc lập quan sát được

Phương pháp: Hồi quy (Regression)

Cách thực hiện bằng hàm:

Tên hàm	Công dụng và cú pháp
SLOPE	<p>Dùng để tính hệ số b của phương trình hồi quy đơn tuyến tính</p> <p>Cú pháp : SLOPE(known_y's, known_x's)</p> <p>Trong đó:</p> <ul style="list-style-type: none">known_y's là các giá trị hoặc vùng địa chỉ của tập số liệu phụ thuộc quan sát được.known_x's là các giá trị hoặc vùng địa chỉ của tập số liệu độc lập quan sát được
CORREL	<p>Dùng để tính hệ số tương quan giữa X và Y</p> <p>Cú pháp : CORREL(array1,array2)</p> <p>Trong đó: array1, array2 tương ứng là tập số liệu phụ thuộc và tập số liệu độc lập quan sát được.</p>

Phương pháp: Hồi quy (Regression)

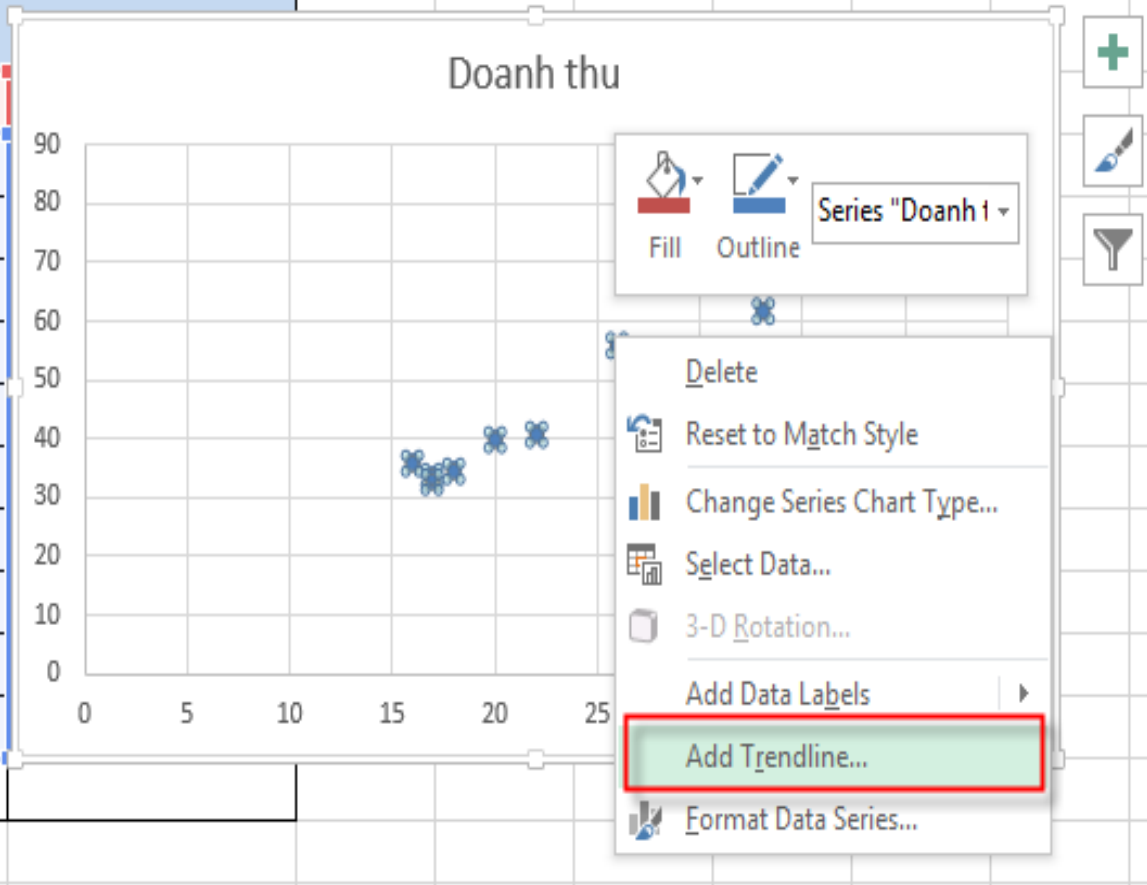
Cách thực hiện bằng đồ thị:

- ▶ Bước 1: Chuẩn bị bảng số dữ liệu cần dự báo
- ▶ Bước 2: Chọn lần lượt vùng địa chỉ chứa biến phụ thuộc Y, và vùng địa chỉ chứa biến độc lập X
- ▶ Bước 3: Vẽ đồ thị dạng **Scatter**
- ▶ Bước 4: Click chuột phải vào data series, chọn **Add Trendline**
- ▶ Bước 5: Tùy chọn hiển thị trong **Trendline Options**
 - ☐ **Linear**: dạng đường thẳng
 - ☐ **Display Equation on chart**
 - ☐ **Display R-squared value on chart**

Phương pháp: Hồi quy (Regression)

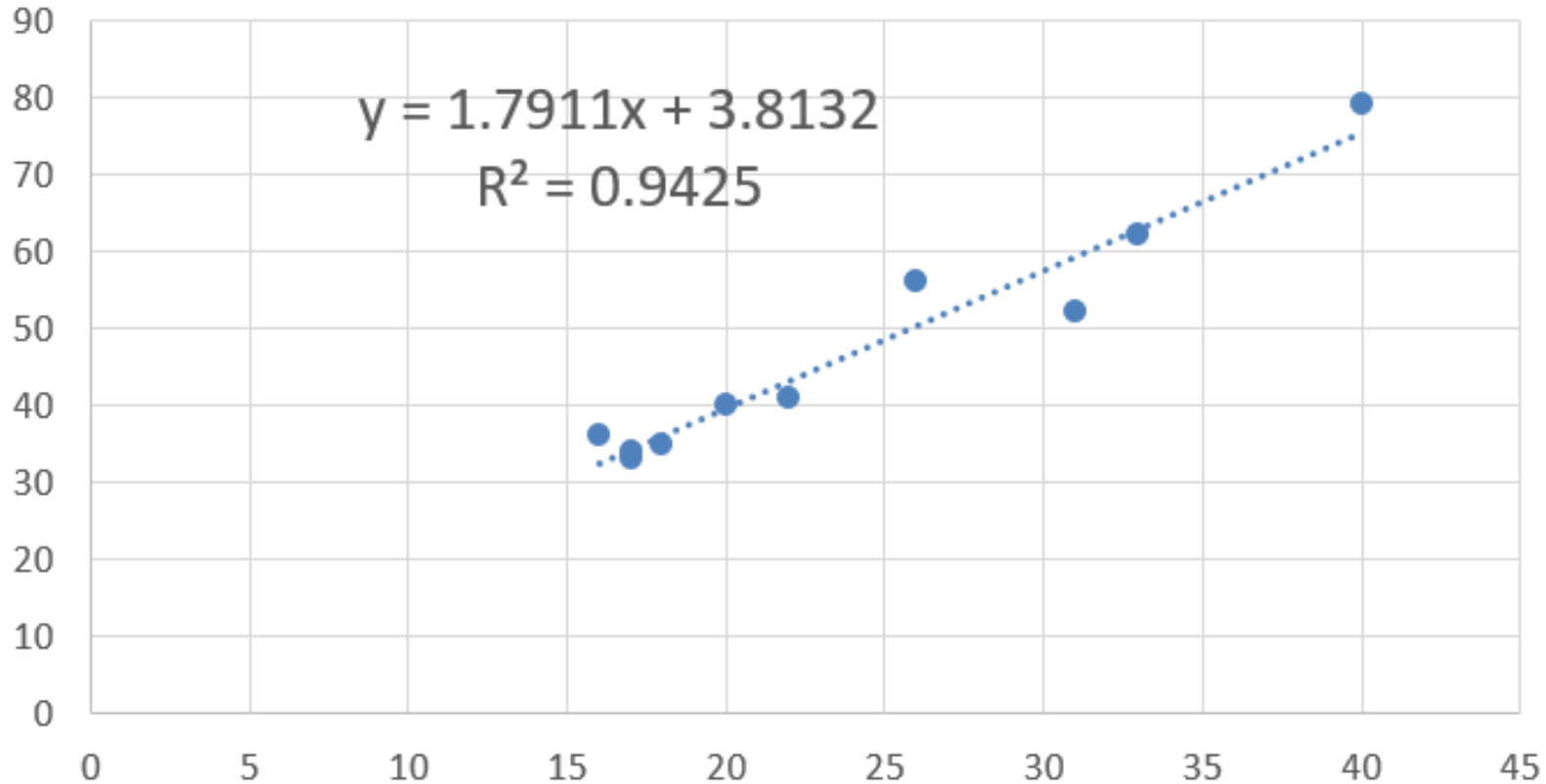
PHÂN TÍCH DOANH THU - CHI PHÍ NĂM 2016

Tháng	Chi phí quảng cáo	Doanh thu
1	20	40
2	17	34
3	22	41
4	31	52
5	16	36
6	40	79
7	33	62
8	18	35
9	17	33
10	26	56
11	35	



Phương pháp: Hồi quy (Regression)

Doanh thu

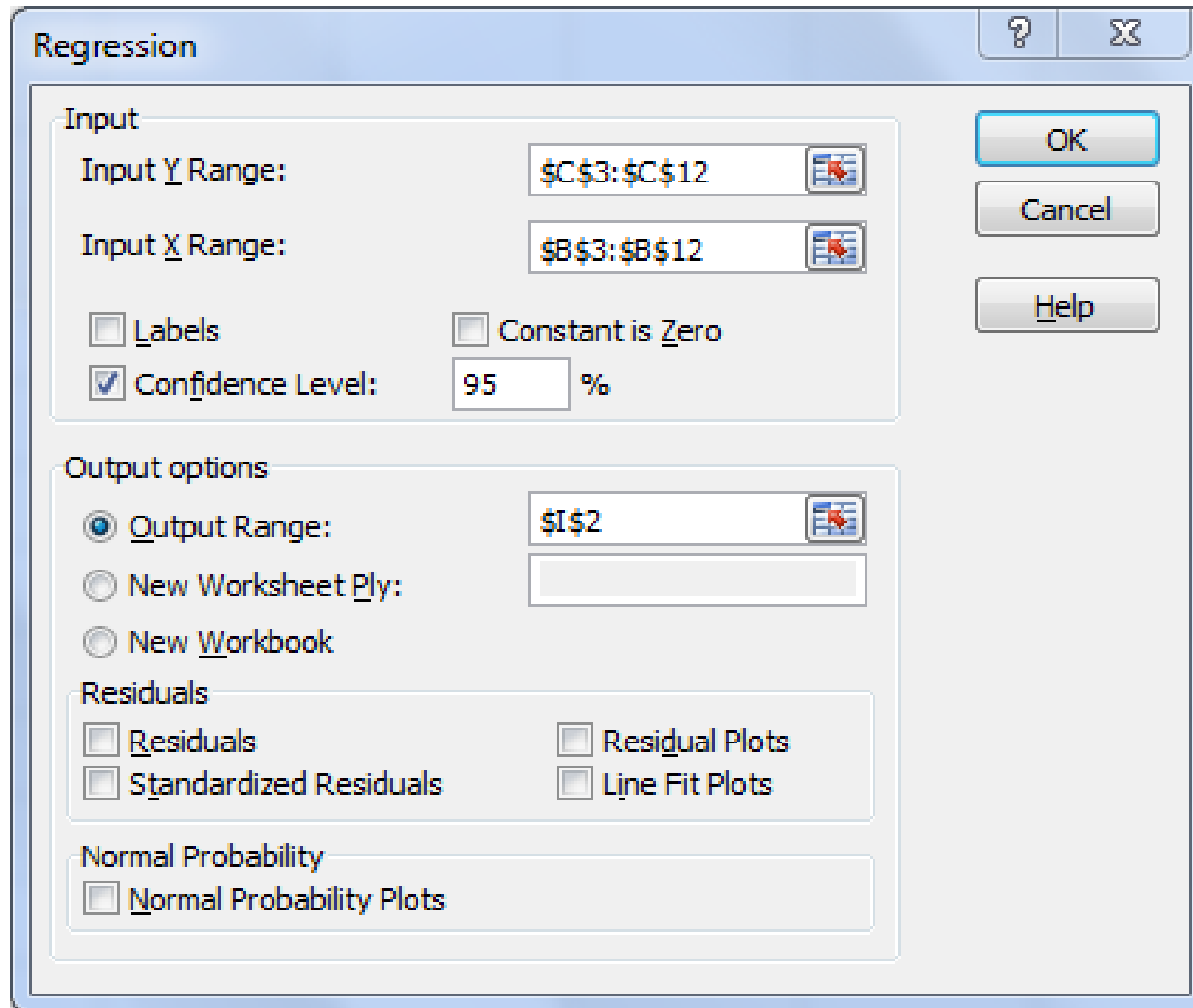


Phương pháp: Hồi quy (Regression)

Cách thực hiện bằng công cụ Regression:

- ▶ **Bước 1:** Chuẩn bị bảng số dữ liệu cần dự báo
- ▶ **Bước 2:** Chọn lệnh **Data** → **Data Analysis** → **Regression**, xuất hiện hộp thoại **Regression**

Phương pháp: Hồi quy (Regression)



Regression

Input

Input Y Range: \$C\$3:\$C\$12

Input X Range: \$B\$3:\$B\$12

☐ Labels ☐ Constant is Zero

☒ Confidence Level: 95 %

Output options

☒ Output Range: \$I\$2

☐ New Worksheet Ply:

☐ New Workbook

Residuals

☐ Residuals ☐ Residual Plots

☐ Standardized Residuals ☐ Line Fit Plots

Normal Probability

☐ Normal Probability Plots

OK

Cancel

Help

Phương pháp: Hồi quy (Regression)

Cách thực hiện bằng công cụ Regression:

- ▶ **Bước 3:** Khai báo các thông số Input và Output Options
 - ❑ **Input Y Range:** Vùng địa chỉ chứa biến phụ thuộc Y
 - ❑ **Input X Range:** Vùng địa chỉ chứa các biến độc lập X (Có thể chọn nhiều biến X trong trường hợp hồi quy đa biến)
 - ❑ **Labels:** Tích vào mục này để khẳng định ô (các ô) đầu tiên không chứa dữ liệu hồi quy
 - ❑ **Constant is Zero:** Tích vào mục này để khẳng định hệ số tự do của hàm hồi quy tuyến tính $a = 0$

Phương pháp: Hồi quy (Regression)

Cách thực hiện bằng công cụ Regression:

- ❑ **Confidentce Level:** Độ tin cậy của hồi quy (mặc định là 95%) bằng $1-\alpha$, với α là mức ý nghĩa hay xác suất mắc sai lầm loại một, bác bỏ H_0 trong khi H_0 đúng.
- ❑ **Output Range:** Vùng hoặc ô phía trên bên trái của vùng chứa kết quả
- ❑ **New Worksheet Ply:** In kết quả ra một sheet khác
- ❑ **New Workbook:** In kết quả ra một file Excel mới
- ❑ **Residuals:** Sai số do ngẫu nhiên
- ❑ **Standardardlized Residuals:** Chuẩn hóa sai số
- ❑ **Residuals Plots:** Đồ thị sai số
- ❑ **Line Fit Plots:** Đồ thị hàm hồi quy tuyến tính
- ❑ **Normal Probability Plots:** Đồ thị xác suất phân phối chuẩn

Phương pháp: Hồi quy (Regression)

SUMMARY OUTPUT								
<i>Regression Statistics</i>								
Multiple R	0.970806962							
R Square	0.942466158							
Adjusted R Square	0.935274428							
Standard Error	3.857979385							
Observations	10							
ANOVA								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
Regression	1	1950.527961	1950.527961	131.0485967	3.0676E-06			
Residual	8	119.0720395	14.88400493					
Total	9	2069.6						
	<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>
Intercept	3.813157895	3.948295423	0.965773197	0.362441329	-5.291627678	12.91794347	-5.291627678	12.91794347
X Variable 1	1.791118421	0.156461725	11.4476459	3.0676E-06	1.430317037	2.151919805	1.430317037	2.151919805

PHÂN TÍCH TỐI ƯU

Khảo sát tình huống

Một nhà quản lý dự án nông nghiệp muốn lựa chọn phương án trồng trọt bao nhiêu tấn lúa mì và lúa gạo để tối đa hóa lợi nhuận của dự án dựa trên các số liệu sau:

Số liệu đầu vào đối với một đơn vị sản phẩm	Loại sản phẩm		Khả năng lớn nhất của các nguồn tài nguyên sẵn có
	Lúa gạo	Lúa mì	
Diện tích đất (ha/tấn)	2	3	50 ha
Lượng nước (m ³ /tấn)	6	4	90 m ³
Nhân công (người/tấn)	15	12	250 người
Lợi nhuận (USD/tấn)	18	21	

Các bước lập mô hình

► Bước 1: Xác định biến quyết định

Gọi x_1 là lượng lúa gạo, x_2 là lượng lúa mì (tấn) cần sản xuất

► Bước 2: Xác định hàm mục tiêu

Mục tiêu bài toán là tối đa hóa lợi nhuận ta có

$$P = P(\text{lúa gạo}) + P(\text{lúa mì}) = 18x_1 + 21x_2 \rightarrow \max$$

► Bước 3: Xác định hệ ràng buộc

□ Ràng buộc tài nguyên sử dụng:

Về diện tích đất: $2x_1 + 3x_2 \leq 50$

Về nước tưới: $6x_1 + 4x_2 \leq 90$

Về lao động: $15x_1 + 12x_2 \leq 250$

□ Ràng buộc tự nhiên: $x_1, x_2 \geq 0$

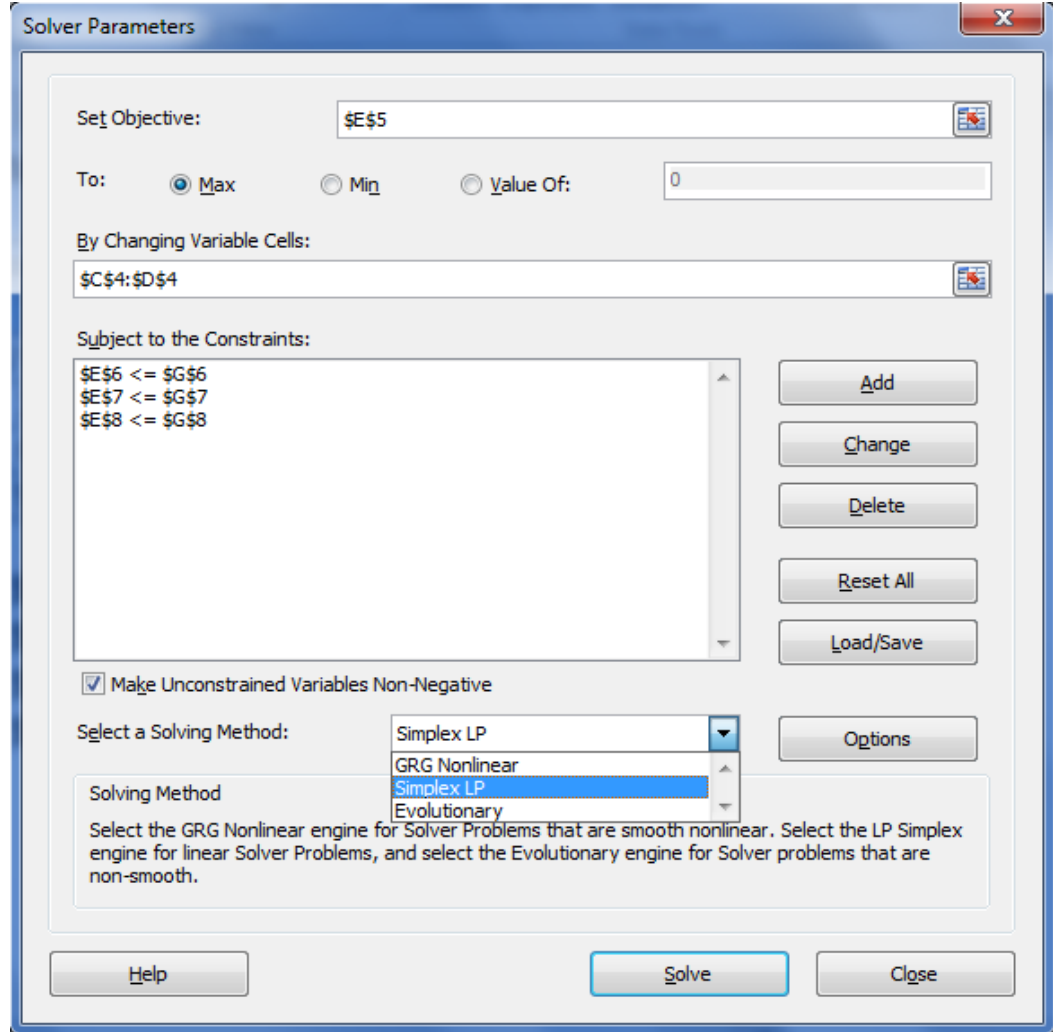
Công cụ SOLVER để giải mô hình kinh tế

Bước 1: Thiết lập bảng tính

	A	B	C	D	E	F	G
1	Lựa chọn phương án sản xuất						
2							
3			Lúa gạo	Lúa mì	Vé trái		Vé phải
4	Biến quyết định	Sản lượng	1	1			
5	Hàm mục tiêu	Lợi nhuận	\$18	\$21	\$39	→	Max
6	Hệ ràng buộc	Diện tích	2	3	5	<=	50
7		Lượng nước	6	4	10	<=	90
8		Nhân công	15	12	27	<=	250
9							
10	E5 = SUMPRODUCT(\$C\$4:\$D\$4,C5:D5)						
11	E6 = SUMPRODUCT(\$C\$4:\$D\$4,C6:D6)						
12	E7 = SUMPRODUCT(\$C\$4:\$D\$4,C7:D7)						
13	E8 = SUMPRODUCT(\$C\$4:\$D\$4,C8:D8)						

Công cụ SOLVER để giải mô hình kinh tế

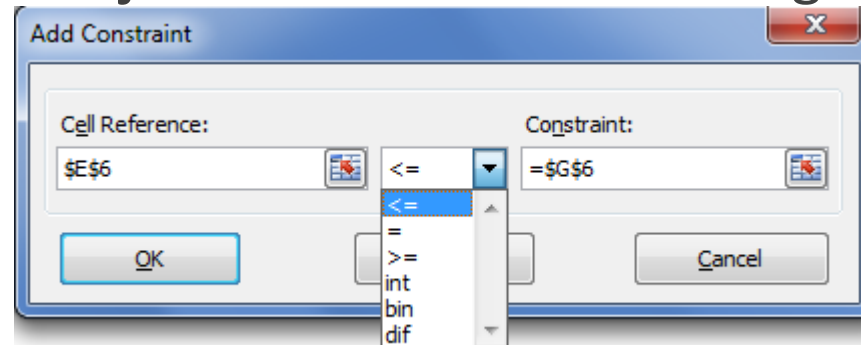
**Bước 2: Chọn lệnh
Data → Analysis →
Solver**



Công cụ SOLVER để giải mô hình kinh tế

► Khai báo các tham số của bài toán:

- ❑ **Set Objective:** Nhập ô chứa hàm mục tiêu, trong trường hợp này là **\$E\$5**.
- ❑ **To:** Chọn **Max** vì bài toán này là tối đa hóa lợi nhuận
- ❑ **By Changing Variable Cells:** Nhập ô chứa các biến quyết định, trong trường hợp này là **\$C\$4 :\$D\$4**
- ❑ Đưa các ràng buộc vào **Subject to the Constraints** bằng cách nhấn nút **Add**



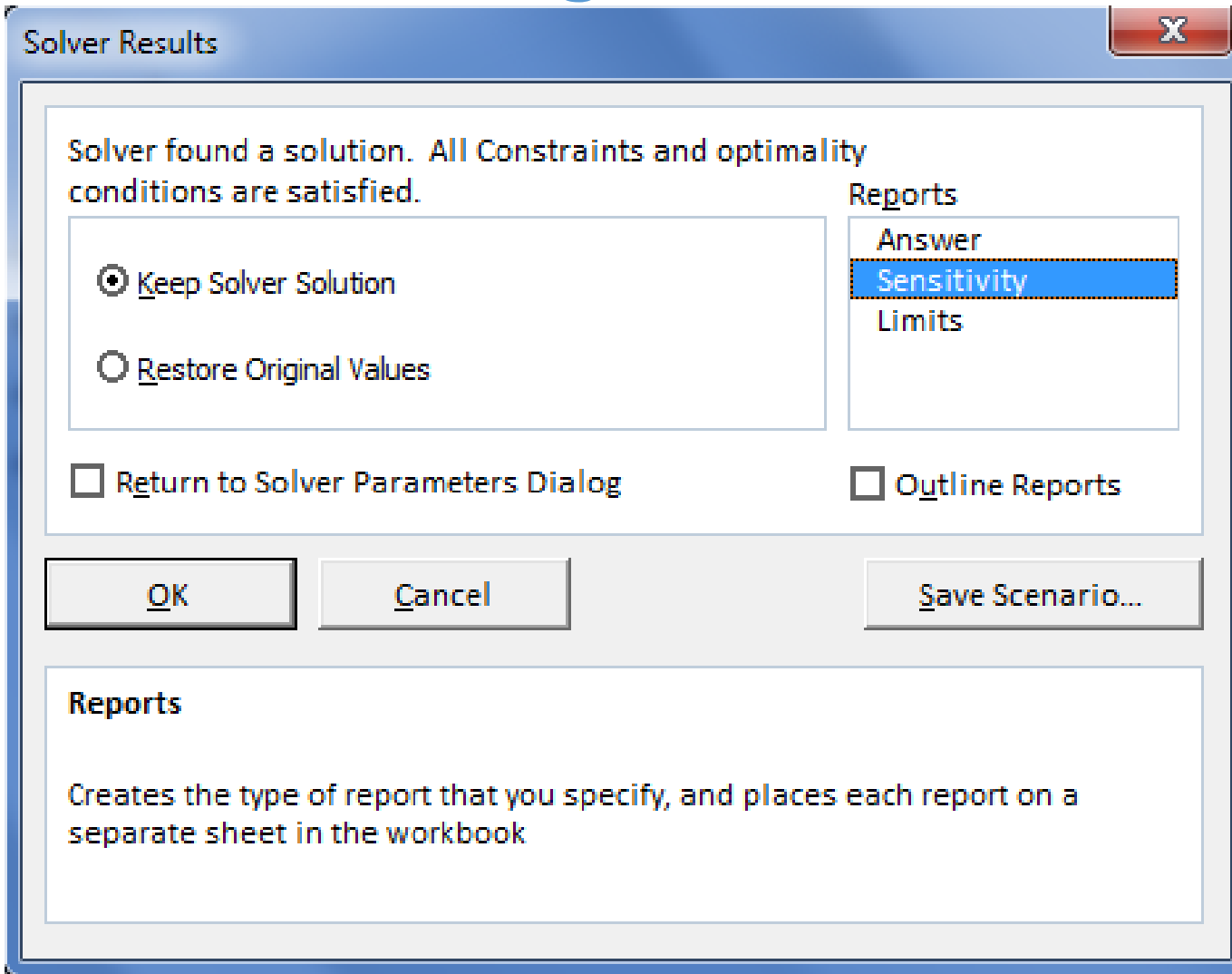
Công cụ SOLVER để giải mô hình kinh tế

Bước 3: Nhấn nút Solve để giải mô hình

Khai báo các lựa chọn trong hộp thoại Solver Results

- ☐ **Keep Solver Solution:** Giữ kết quả và in ra bảng tính.
- ☐ **Restore Original Values:** Huỷ kết quả vừa tìm được và trả các biến về tình trạng ban đầu.
- ☐ **Save Scenario:** Lưu kết quả vừa tìm được thành một tình huống để có thể xem lại sau này.
- ☐ Có thể xuất hiện thêm các dạng báo cáo trong kết quả: Answer, Sensitivity và Limits

Công cụ SOLVER để giải mô hình kinh tế



Công cụ SOLVER để giải mô hình kinh tế

Bước 4: Nhấn nút OK để xem kết quả

	A	B	C	D	E	F	G
1	Lựa chọn phương án sản xuất						
2							
3			Lúa gạo	Lúa mì	Vé trái		Vé phải
4	Biến quyết định	Sản lượng	7	12			
5	Hàm mục tiêu	Lợi nhuận	\$18	\$21	\$378	→	Max
6	Hệ ràng buộc	Diện tích	2	3	50	<=	50
7		Lượng nước	6	4	90	<=	90
8		Nhân công	15	12	249	<=	250
9							
10	E5 = SUMPRODUCT(\$C\$4:\$D\$4,C5:D5)						
11	E6 = SUMPRODUCT(\$C\$4:\$D\$4,C6:D6)						
12	E7 = SUMPRODUCT(\$C\$4:\$D\$4,C7:D7)						
13	E8 = SUMPRODUCT(\$C\$4:\$D\$4,C8:D8)						

Công cụ SOLVER để giải mô hình kinh tế

Các báo cáo kèm theo lời giải

	A	B	C	D	E	F	G	H
1	Microsoft Excel 14.0 Sensitivity Report							
2	Report Created: 25/03/2015 15:38:10							
3								
4								
5	Variable Cells							
6			Final	Reduced	Objective	Allowable	Allowable	
7	Cell	Name	Value	Cost	Coefficient	Increase	Decrease	
8	\$C\$4	Sản lượng Lúa gạo	7	0	18	13.5	4	
9	\$D\$4	Sản lượng Lúa mì	12	0	21	6	9	
10								
11	Constraints							
12			Final	Shadow	Constraint	Allowable	Allowable	
13	Cell	Name	Value	Price	R.H. Side	Increase	Decrease	
14	\$E\$6	Diện tích Vế trái	50	5.4	50	0.833333333	20	
15	\$E\$7	Lượng nước Vế trái	90	1.2	90	0.476190476	23.33333333	
16	\$E\$8	Nhân công Vế trái	249	0	250	1E+30	1	

Hỏi đáp

