

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC NÔNG NGHIỆP HÀ NỘI**

**PGS.TS. NGÔ THỊ THUẬN (Chủ biên)
THS. LÊ VĂN BỘ, THS. LÊ NGỌC HƯỚNG
THS. NGUYỄN THI NHUẬN**

GIÁO TRÌNH TIN HỌC ỨNG DỤNG

HÀ NỘI – 2007

Chương I. NHẬP MÔN TIN HỌC ỨNG DỤNG

1. Đối tượng nghiên cứu của môn học

1.1 Một số khái niệm

a) Máy vi tính và sự hình thành máy vi tính

- Máy vi tính là một máy tính nhỏ được John V. Blann Kenbaker chế tạo năm 1971, đây là một chiếc máy tính Kenbak được chế tạo đầu tiên với kích thước nhỏ. Máy Kenbak chế tạo từ 130 mạch tổ hợp IC nhưng thực chất nó chỉ là 1 máy tính cỡ nhỏ. Nó có thể thực hiện 1000 lệnh/giây, giá 750 USD (thời điểm 1971)

- Nếu coi máy vi tính là một máy tính nhỏ sử dụng bộ vi xử lý phổ biến hiện nay thì do một người Việt quốc tịch Pháp chế tạo ra vào năm 1973, chỉ 6 tháng sau khi xuất hiện bộ vi xử lý. Đó là ông Andre Trương Ngọc Thi và chiếc máy Micral của hãng R3E.

- Năm 1975, một chiếc máy khác cũng sử dụng bộ vi xử lý, đó là chiếc Altair 8.800 của một hãng Mỹ chế tạo.

- Năm 1985, nhà bảo tàng máy tính ở Boston đã tìm hiểu xem ai là người phát minh ra máy vi tính đầu tiên nhưng không xác định được một cách chính xác. Vì vậy, giải thưởng được trao cho 3 máy tính nói trên.

b) Tin học

Nước Pháp là nơi đầu tiên đặt tên gọi cho ngành tin học, người ta định nghĩa: *“Tin học là tập hợp các khoa học, kỹ thuật về việc xử lý tự động và hợp lý thông tin mang kiến thức của con người để lưu giữ và truyền gửi chúng”*.

Từ điển Bách khoa 1986 của Liên Xô (cũ) định nghĩa: *“Tin học là một lĩnh vực khoa học nghiên cứu cấu trúc và tính chất của thông tin khoa học cùng với việc thu thập, xử lý, lưu giữ, biến đổi và truyền chúng”*.

Tuy định nghĩa có thể khác nhau nhưng nội dung của các định nghĩa ấy đều thống nhất ở một điểm: *Tin học là khoa học chuyên nghiên cứu về một đối tượng cụ thể, đó là thông tin và dữ liệu. Công cụ không thể thiếu đối với người nghiên cứu tin học là máy tính và các ngôn ngữ giao tiếp*.

Như vậy, tin học là một ngành khoa học công nghệ, nghiên cứu các quá trình xử lý thông tin một cách tự động với sự trợ giúp của máy tính điện tử.

Infomatic = Information + Automatic = Computer Science

c) Tin học ứng dụng

Tin học ứng dụng là một môn khoa học thuộc công nghệ thông tin nhằm ứng dụng tin học vào cuộc sống. Tùy từng lĩnh vực mà tin học được áp dụng khác nhau.

d) Tin học quản lý

Tin học quản lý là một lĩnh vực của tin học ứng dụng nhằm sử dụng các phương pháp quản lý thông tin của tin học vào lĩnh vực quản lý kinh tế, xã hội, hành chính, giáo dục.v.v..

e) Tin học ứng dụng trong kinh tế và quản trị kinh doanh

Tin học ứng dụng trong kinh tế và quản trị kinh doanh là một bộ phận của tin học quản lý nhằm ứng dụng các phương pháp quản lý thông tin của tin học trong quản lý nền kinh tế và các loại hình doanh nghiệp.

1.2 Đối tượng, phạm vi ứng dụng tin học trong kinh tế và quản trị kinh doanh

Đối tượng nghiên cứu của tin học ứng dụng trong kinh tế và quản trị kinh doanh là toàn bộ dữ liệu nghiên cứu, các hàm, các lệnh, công thức tính, các kỹ thuật đồ họa, cơ sở dữ liệu, các kỹ thuật tổng hợp, phân tích, xây dựng mô hình kinh tế, lập kế hoạch sản xuất kinh doanh, dự báo kết quả hoạt động sản xuất kinh doanh và căn cứ đề ra các quyết định lựa chọn tối ưu.

Phạm vi nghiên cứu của môn học: Nghiên cứu ứng dụng tổng hợp, xử lý, phân tích dữ liệu về các lĩnh vực kinh tế, kế hoạch, kế toán quản trị, tài chính, quản trị kinh doanh, dự án, đầu tư, ngân hàng, thống kê, kinh tế lượng, chính sách, v.v...

1.3 Phương pháp, kỹ thuật ứng dụng tin học trong kinh tế - quản trị kinh doanh

- Các phương pháp Thống kê - Kinh tế lượng - Toán kinh tế
- Các phương pháp Kế toán - Tài chính - Dự án - Đầu tư - Ngân hàng
- Các phương pháp phân tích của Vi mô - Vĩ mô - Chính sách
- Phần mềm EXCEL, SPSS, v.v... được sử dụng làm công cụ kỹ thuật ứng dụng tin học kinh tế và quản trị kinh doanh.

2. Phần mềm ứng dụng trong kinh tế và quản trị kinh doanh

2.1 Các phần mềm chung

Hiện nay có rất nhiều phần mềm ứng dụng trong kinh tế, tùy thuộc vào khả năng trang thiết bị hiện có của cơ sở đào tạo cũng như trình độ tin học, trình độ chuyên môn và khả năng vận dụng linh hoạt của người nghiên cứu mà sử dụng các chương trình ứng dụng sau:

- EXCEL: Bảng tính, phân tích khái quát nguồn số liệu.
- FOXPRO: Quản lý dữ liệu.
- ACCESS: Bảng tính, phân tích số liệu tổng quát, quản lý dữ liệu.
- MFIT, TSP, EVIEW: Phần mềm chuyên dụng cho kinh tế lượng.
- IRISTAF, STAT, STATA, MINITAB: Phần mềm chuyên dụng cho thống kê sinh học.
- SPSS, SAS: Phần mềm chuyên dụng cho thống kê kinh tế xã hội và kinh tế lượng.
- SHAZAM, LIMDEP: Phần mềm chuyên dụng cho, toán kinh tế, kinh tế lượng, v.v...
- BESTFIT, RISK: Phần mềm chuyên dụng cho bảo hiểm và rủi ro.
- IPS: Phần mềm tài chính
- LINDO: Phần mềm quy hoạch tuyến tính.
- POM: Phần mềm quy hoạch tuyến tính, thống kê, kinh tế lượng.
- KTM: Phần mềm kế toán máy, v.v....

2.2 Phần mềm Excel

EXCEL là bảng tính điện tử được thiết kế hoạt động trong môi trường WINDOWS. Hiện nay, EXCEL là bảng tính tiện ích, nó tương thích và có thể truy nhập dữ liệu từ các phần mềm bảng tính khác như LOTUS, QUATTRO hoặc từ các chương trình khác như FOXPRO, ACCESS, SPSS, STATA, MINITAB, v.v... Ngoài tính năng tính toán như bảng tính thông thường, EXCEL còn có thể sử dụng để quản lý cơ sở dữ liệu, thiết kế đồ họa, xử lý một số chương trình thống kê phức tạp như tương quan, hồi quy, bài toán quy hoạch tuyến tính, T-TEST, ANOVA v.v...

EXCEL rất phù hợp với sinh viên ở nhiều trình độ khác nhau, từ bình thường đến chuyên sâu. So với các phần mềm khác thì EXCEL rất dễ học. Tuy nhiên, việc học phần mềm EXCEL ứng dụng vào kinh tế đòi hỏi phải có thời gian và sự nỗ lực lớn của người học. Nếu chúng ta chịu khó tìm tòi, thực tập nhiều trên máy và nếu những điều chúng ta đem áp dụng trên máy lại vượt quá những gì đã được trình bày ở bài giảng này thì sự cố gắng và đầu tư của chúng ta thực sự có nhiều giá trị. Các chức năng cơ bản của EXCEL bao gồm:

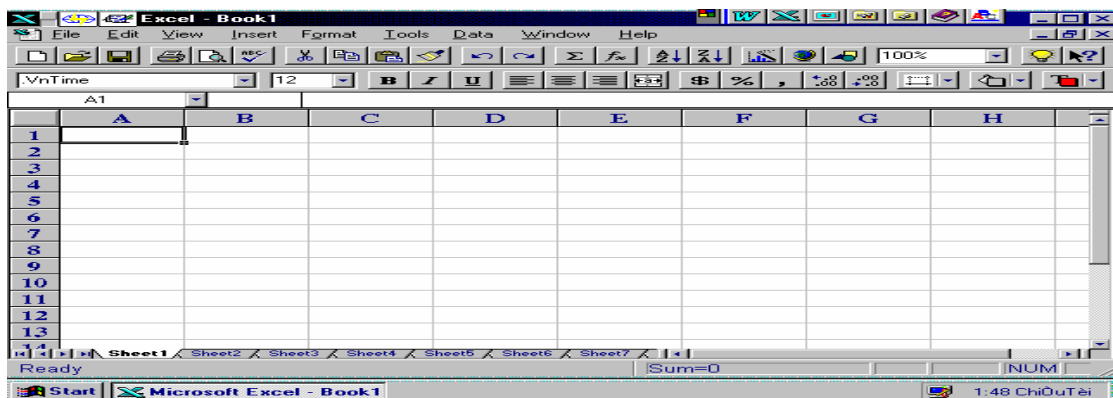
- Tự động điền một chuỗi số tuần tự hoặc một dãy dữ liệu (đã khai báo trước) vào những ô trong phạm vi của bảng tính do người sử dụng quy định (dùng tính năng Auto Fill hoặc Fill Handle).

- Di chuyển, sao chép dữ liệu trong bảng tính được thực hiện nhanh chóng và đơn giản nhờ kỹ thuật rê chuột (Drag and Drop).
- Tập hợp những bảng tính có quan hệ với nhau vào trong cùng một tập tin (Workbook).
- Tính toán dựa trên các hàm số có sẵn của EXCEL hoặc sử dụng công thức do người sử dụng thiết lập.
- Dễ dàng theo dõi sự phụ thuộc của các ô dữ liệu trong bảng tính (Auditing).
- Phân loại, tổng hợp các nhóm dữ liệu trong bảng tính (Subtotals).
- Nhập và liên kết dữ liệu trong nhiều bảng tính (sử dụng Consolidate).
- Chọn lọc nhanh chóng các dữ liệu theo điều kiện do người sử dụng ấn định (dùng Data Filter).
- Phân tích và hệ thống dữ liệu theo những cách trình bày khác nhau (Pivot Table).
- Trao đổi và liên kết dữ liệu giữa các chương trình ứng dụng khác nhau.
- Dựng các đồ thị, biểu đồ minh họa cho các số liệu của bảng tính, v.v...

a) Giới thiệu bảng tính EXCEL

Đang ở Windows ta nhấp kép vào biểu tượng của EXCEL (biểu tượng này thường nằm trong cửa sổ Microsoft office hay trên thanh biểu tượng của office) ta thấy mở ra cửa sổ EXCEL với thiết kế tương tự cửa sổ Winword:

- **Thanh tiêu đề** nằm trên cùng, bao gồm: Nút Control ở bên trái và các nút mở to, thu nhỏ, nút tắt ở bên phải, giữa là tên tệp EXCEL đang mở. Trong EXCEL, mỗi tệp được gọi là 1 quyển sách (Book) có nhiều trang (Worksheet).
- **Thanh Menu** (Menu Bar) ghi các Menu: File, Edit, View, Insert, Format, Tools, Data, Windows, Help. Khi nhấp vào một Menu sẽ được một cửa sổ đọc với các Menu
- **Thanh công cụ** (Tool Bar) với các biểu tượng mà một số hình thức và tác dụng giống y như trong Winword: New, Open, Save, Print, Preview, Cut, Copy, Paste, Format, Undo, ... Ngoài ra còn có các biểu tượng mới riêng cho EXCEL như SUM tự động Σ , Hàm định sẵn f_x , Sort, đồ thị Chart, khung để viết văn bản.
- **Thanh định dạng** (Formatting bar) với một số biểu tượng và nội dung như bên Winword như Font chữ, kích thước chữ, in đậm B, in nghiêng I, ..., ngoài ra còn có thêm các biểu tượng riêng của EXCEL như tăng số thập phân, giảm số thập phân, chọn màu nền, chọn màu chữ,...
- **Thanh Formula:** Khoảng đầu ghi địa chỉ ô hiện hành. Khoảng giữa có 3 ô ghi dấu bỏ (X màu đỏ), chấp nhận (V màu xanh lam) và chọn hàm (f_x). Khoảng bên phải hiện nội dung của ô khi ta nhập dữ liệu hay khi cần xem lại nội dung của ô. Phần giữa là trang tính hiện hành, bên phải là thanh trượt dọc để kéo dọc màn hình lên xuống.
- **Trang hiện hành** bao gồm:



- Hàng trên cùng là tên của 256 cột, có tên lần lượt là A, B, ..., AA, AB, ..., IV.

- Cột đầu là tên các hàng có tên lần lượt là 1 cho đến 65536.
- Giao của hàng và cột tạo thành các ô, mỗi ô có 1 địa chỉ hay còn gọi là thanh biến. Địa chỉ của ô (miền, vùng) có các dạng sau:
 - + Địa chỉ tương đối ghi dưới dạng: cột dòng, ví dụ A3 (cộtAdòng3)
 - + Địa chỉ tuyệt đối ghi dưới dạng: \$cột\$dòng, ví dụ \$A\$3 (\$cột\$dòng3)
 - + Địa chỉ hỗn hợp ghi dưới dạng: \$cột_dòng hoặc cột\$dòng, ví dụ: \$A3; A\$3.
- Một nhóm ô liên nhau tạo thành một chữ nhật được ghi địa chỉ bằng địa chỉ góc trên bên trái : địa chỉ góc dưới bên phải, thí dụ: A1:C5, hay được tham chiếu bằng một tên (Name).
- Một nhóm nhiều miền rời nhau được ghi bằng cách ghi liên tiếp địa chỉ các miền cách nhau bằng dấu phẩy (,) thí dụ: A3:C9, D5:H20, I9:K30,....
- **Con trỏ** (4 phím mũi tên) chỉ có thể di chuyển trong vùng bảng tính; còn **con chuột** di chuyển trong vùng bảng tính có biểu tượng là dấu “cộng”, trên các thanh công cụ, thực đơn là biểu tượng mũi tên.
- **Dưới cùng ghi tên các trang** (Worksheet) thường có 3 trang ghi từ Sheet1 đến Sheet3. Bên trái có các nút chọn trang đầu, chọn trang trước, chọn trang sau, chọn trang sau cùng. Bên phải là thanh trượt ngang để kéo màn hình.

b) Các nhóm lệnh cơ bản

- **File:** Nháy vào File (Alt+F) ta được cửa sổ đọc bao gồm các thực đơn (Menu): New (mở tệp mới), Open (mở tệp cũ), Close (đóng tệp), Save (cất tệp), chú ý lần đầu phải đặt tên tệp với đuôi mặc nhận là XLS và có thể ghi vào chỗ ta muốn bằng cách chọn ổ đĩa và thư mục thích hợp. Save As (ghi với tên khác), Page setup để định trang in, Print để in v.v... Phía dưới là tên của tệp vừa mở và cuối cùng là Exit để thoát khỏi EXCEL.
- **Edit:** Nháy vào Edit (Alt+D) ta được cửa sổ đọc bao gồm các thực đơn: Undo (bỏ lệnh vừa thực hiện), Cut (cắt), Paste (dán), Paste Special (Dán đặc biệt), Clear (xoá với khả năng xoá tất cả (All), xoá nội dung (Contents), xoá định dạng (Format), xoá chú thích (Notes), Bỏ (Delete), Bỏ cả trang (Delete Sheet) v.v... Một số các Menu con ít dùng vì đã có các biểu tượng với tính năng tương đương trên các thanh công cụ như Cut, Copy, Paste,...
- **View:** Menu này cho chúng ta biết đang mở các thanh nào (Standard bar, Formatting bar, Formula bar) và muốn mở thêm một số thanh công cụ thì vào Tools bar để mở thêm, chế độ toàn màn hình (Full screen) hoặc phóng to (Zoom)...
- **Insert:** Ngoài các Menu con như chèn ô, chèn hàng, chèn cột còn có thể chèn các tài liệu từ các nguồn khác, tạo tên miền...
- **Format:** Trong menu này có Menu con Cell để tạo khuôn cho các ô ta chọn, nếu muốn tạo khuôn cho cả trang thì chọn Menu con Style sau đó có thể chọn kiểu chữ (Font), kích thước (Size), màu nền, màu chữ, cách viết ngày tháng (Date), cách viết số (Number).
- **Tool:** Menu này có các phần Add-Ins để nhập thêm một số phần như Solver, Data analysis, View manager...

Khi đã có các phần đó thì trên Menu mới có các Menu con Solver và Data analysis mà chúng ta sẽ gặp trong các phần sau. Ngoài ra còn có Macro (vĩ lệnh) để viết một số lệnh và chương trình con.

- **Data:** Menu này chuyên về cơ sở dữ liệu với nhiều mục mà chúng ta sẽ trình bày ở phần sau như Table (Bảng), Sort (sắp xếp thứ tự), Filter (lọc), Pivot Table (bảng tổng kết).
- **Windows:** Menu điều khiển mở đóng, theo dõi các cửa sổ tương tự như Winword.
- **Help:** Menu này gọi trợ giúp và cấu tạo như phần Help của Winword.

c) Cách nhập các dữ liệu

Vào EXCEL xong ta mở tệp mới bằng cách chọn New hoặc mở tệp cũ (Open) và bước vào trang 1 (Worksheet1).

- Đặt tên cho trang: Để chuột vào ô Sheet1 (phía dưới màn hình) sau đó nhấp chuột phải và chọn Menu Rename để đổi sang tên mới thí dụ Sổ liệu 1 hay Bảng 1,...
- Có thể mở rộng hay thu hẹp cột bằng cách đặt chuột vào bờ ngăn cách 2 tên cột, khi nào chuột chuyển sang dạng chữ thập kéo theo 1 cột các chấm mờ thì kéo chuột sang phải để mở rộng hay sang trái để thu hẹp cột, nếu trong các ô đã có dữ liệu thì nhấp kép tại bờ sẽ mở rộng cột đến mức có thể nhìn thấy toàn bộ dữ liệu trong ô chứa dữ liệu dài nhất. Làm tương tự với hàng bằng cách đặt chuột vào bờ ngăn cách 2 tên hàng.

* Chọn miền làm việc

- Chọn 1 ô bằng cách nhấp vào ô đó, ô đó sẽ đổi màu. Chọn 1 cột bằng cách nhấp vào tên cột, chọn 1 hàng bằng cách nhấp vào tên hàng.
- Chọn 1 miền bằng cách nhấp chuột vào ô đầu (ô trên cùng bên trái) sau đó rê sang phải và xuống dưới cho đến hết miền (ô dưới cùng bên phải).
- Có thể chọn nhiều miền không kế tiếp bằng cách chọn xong miền 1 ấn và giữ Ctrl rồi chọn miền miền tiếp theo.

* Nhập dữ liệu

- Khi chọn 1 ô xong ta nhập dữ liệu, nhập xong thì gõ Enter hay dùng các phím mũi tên dịch chuyển khỏi ô, hoặc dùng chuột nhấp vào ô chấp nhận trên thanh Formula.
- Nhập số: Nói chung, số được giải phóng phải trừ trường hợp chúng ta chọn lại cách giống khác bằng các biểu tượng giống lẻ. Song hiện lên như thế nào tùy thuộc vào cách chọn dạng Number trong Style. Có thể nhập phân số nhưng chú ý sau khi ghi phần nguyên xong phải cách 1 khoảng trống rồi mới ghi sang phân số (Thí dụ: 9 4/5) nếu thấy chưa đủ hay nhiều số thập phân quá thì có thể thêm, bớt bằng cách kích chuột vào biểu tượng trên thanh công cụ.
- Nhập văn bản: Khi gõ ký tự không phải số thì EXCEL nhận ra đó là nhãn (Label) hay còn gọi là văn bản (Text) và ghi kết quả theo cách giống trái, trừ trường hợp chúng ta chọn lại cách giống bằng các biểu tượng giống lẻ hay gõ các ký hiệu (*) trước văn bản để giống trái, (**) để giống phải, (^) để giống giữa. Khi nhập văn bản nếu dài quá thì EXCEL vẫn nhớ đủ và khi hiện lên có thể cho hiện sang cả các ô bên cạnh nhưng toàn bộ văn bản chỉ thuộc 1 ô.
- Nhập ngày tháng: Nhập dạng số cách nhau / hay - . Tùy thuộc Style mà ta thấy kết quả hiện lên dạng số (4/5/99) hay dạng đan xen (5- April-99).
- Nếu có cùng một nội dung mà cần nhập vào nhiều ô thì có thể nhập vào 1 ô, chọn ô đó rồi đặt chuột vào góc dưới bên phải của ô (nơi có chữ thập đen) để kéo cho đến hết miền cần điền số liệu, đó là cách Drag hay còn gọi là AutoFill.
- Nhập công thức: Khi cần tính toán ta nhập công thức bằng cách ghi = tiếp theo là công thức với các toán hạng là địa chỉ các ô hay miền (các tham biến), các phép tính ở đây là các phép cộng (+), trừ (-), nhân (*), chia (/), lũy thừa (^) hay các phép tính Logic như Và (And) Hoặc (Or), phép tính điều kiện (If).
- Nếu dùng các hàm định sẵn thì không cần mở đầu bằng dấu = mà nhấp ngay vào biểu tượng f_x để chọn hàm, sau đó dựa vào hộp thoại để đưa đối số vào hàm.

Sau khi nhập số liệu có thể chọn một miền rồi tạo dạng (Format) cho nó bằng cách chọn kiểu chữ, kích thước, màu nền, màu chữ nhờ các biểu tượng trên thanh Formula.

* Xoá, cắt, chép, dán

Muốn xoá nội dung 1 ô thì gõ vào phím Del, nếu ô đã chỉ định dạng (chọn kiểu chữ, màu nền, màu chữ) thì phải vào Menu Edit chọn Clear sau đó tùy chọn xoá hết, xoá tương tự như chùi sạch mà không động chạm gì đến các ô khác.

Nếu dùng Menu Edit và chọn Delete thì trước khi xoá EXCEL sẽ hỏi: Sau khi xoá xong sẽ đẩy các ô dưới lên (Shift cell up) hay đẩy các ô bên phải sang (Shift cell left). Như vậy Delete ảnh hưởng đến các ô xung quanh và nếu không cẩn thận thì các kết quả cũ có thể bị xô lệch hàng, lệch cột.

Nếu muốn thêm 1 cột thì nháy vào tên cột nơi định chèn sau đó vào Insert/Column, cột mới sẽ được chèn vào trước cột đã chọn.

Tương tự nếu muốn chèn 1 hàng thì chọn hàng nơi định chèn vào Insert/Row, cột mới được chèn vào trên cột đã chọn.

Nếu muốn thêm 1 ô thì chọn ô nơi định chèn vào Insert/ Cell, máy sẽ hỏi Shift cells right, tức là đẩy ô cũ sang phải, hay Shift cells down, tức là đẩy ô cũ xuống dưới. Việc đẩy này sẽ ảnh hưởng đến các ô bên phải hay các ô bên dưới, do đó phải cẩn thận khi dùng.

Nếu muốn cắt 1 ô để rồi chép sang 1 ô khác (rời chỗ) thì có thể chọn ô định cắt, nháy vào biểu tượng Cut, sau đó đến ô mới, nháy vào Paste.

Nếu muốn chép 1 ô sang 1 ô mới thì chọn ô cũ, nháy vào Copy, đến ô mới nháy vào Paste.

Trường hợp chỉ muốn chép lại nội dung (Values) hay chỉ chép lại khuôn dạng (Format) hay cần chuyển cột thành hàng và nhiều cách dán khác thì dùng Paste Special trong menu Edit.

Có thể rời nhanh một miền bằng cách chọn miền rồi đưa chuột vào cạnh miền, sau đó kéo (Drag) và thả sang chỗ khác.

Có thể điền nhanh một dãy số bằng cách điền 2 ô đầu, chọn cả 2 ô, đưa chuột vào góc dưới bên phải miền, chỗ có chữ thập (Handle), rồi kéo xuống dưới hoặc sang ngang, tùy theo 2 ô đầu ghi nội dung gì mà EXCEL suy ra các nội dung tiếp theo căn cứ vào một số quy tắc mà chúng ta có thể tìm hiểu trong menu Edit phần Fill.

* Một số vấn đề về định dạng

Ta có thể định dạng riêng 1 ô hay 1 miền bằng cách chọn ô hay miền rồi vào Format/Cell hoặc có thể định dạng cả Sheet bằng Style sau đó lần lượt lựa chọn các nút để xác định: Font (kiểu chữ), Size (kích thước), in đậm, in nghiêng, có chân, có thể chọn kiểu in ngang hay in dọc, giống trái, giống phải, giống giữa bằng Aligment, chọn đường viền cho Border sau đó có thể lưu định dạng đó để dùng về sau.

* Đặt tên

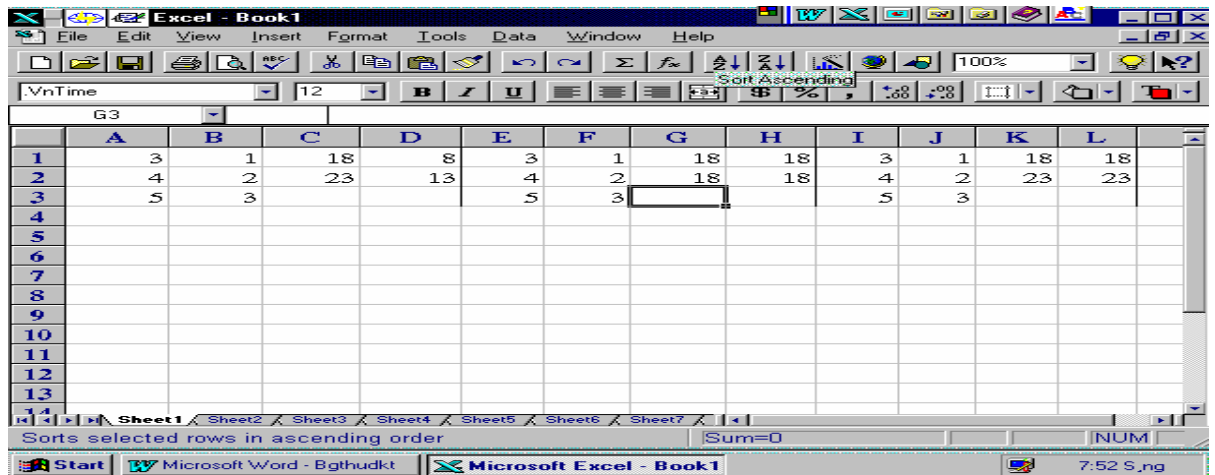
Để tiện sử dụng có thể chọn đặt tên cho 1 miền bằng cách chọn một miền bằng cách chọn miền rồi vào Insert/ Names sau đó đặt tên mới hoặc sử dụng một tên đã có sẵn. Về sau này khi cần tham chiếu đến miền đó ta chỉ việc điền tên thay cho phải kê khai địa chỉ.

d) Cách khai thác công thức, hàm và lệnh

* Các công thức đơn giản

Nhập vào 1 ô đầu = tiếp theo là công thức tính với các toán hạng là các địa chỉ của các ô (tham chiếu) và các toán tử là các phép tính + - * / ^ %.

Cần lưu ý khi sao chép công thức sang các ô khác phải chú ý đến sự thay đổi tham biến nếu trong công thức sử dụng địa chỉ tương đối, tuyệt đối, hỗn hợp:



* Công thức sử dụng địa chỉ tương đối

Nhập vào $C1=2*A1+3*A2$ kết quả được $2*3+3*4=18$

Nếu chép nội dung C1 sang C2 thì được $2*A2+3*A3$ bằng 23

Nếu chép công thức ở C1 sang D1 thì ta được $2*B1+3*B2$ bằng 8

Nếu chép công thức ở C1 sang D2 thì ta được $2*B2+3*B3$ bằng 13

Công thức sử dụng địa chỉ tương đối khi chép nội dung C1 sang C2 thì các địa chỉ của toán hạng sẽ thay đổi theo, A1 thành A2, A2 thành A3 (do sự thay đổi vị trí công thức từ C1 sang C2). Khi chép nội dung C1 sang D1 thì A1 thành B1, A2 thành B2 (do thay đổi công thức từ C1 sang D1). Cuối cùng A1 thành B2, A2 thành B3 khi thay đổi công thức từ C1 sang D2.

* Công thức sử dụng địa chỉ tuyệt đối

Kết quả không hề thay đổi khi chúng ta chép nội dung từ vị trí này sang vị trí khác vì không hề thay đổi tham chiếu.

Nhập vào $G1=2*\$E\$1+3*\$E\2 kết quả $2*3+3*4=18$

Chép sang G2, H1, H2 ta vẫn được 18.

* Công thức sử dụng địa chỉ hỗn hợp

Nhập $K1=2*\$I1+3*\$I2$ kết quả được $2*3+3*4$ bằng 18

Chép sang K2 ta được $2*\$I2+3*\$I3$ bằng 23

Chép sang L1 ta được $2*\$I1+3*\$I2$ bằng 18

Chép sang L2 ta được $2*\$I2+3*\$I3$ bằng 23

Khi viết các công thức phức tạp phải hết sức thận trọng và khi gặp thông báo Error in formula thì phải kiểm tra kỹ các ký tự, các vòng ngoặc, nhiều khi phải xóa đi viết lại.

* Sử dụng các hàm định sẵn

Trong EXCELL có để sẵn khá nhiều hàm và chia ra nhiều loại (Categories). Nháy vào f_x trên thanh Formula hay trên thanh Tools ta vào hộp với 2 nửa:

- Nửa trái là Categories gồm: Most Recently used (Vừa mới dùng), All (Tất cả các hàm), Statistical (Thống kê), Date and Time (Thời gian), Financial (Tài chính), Text (Văn bản), Logical (Logic), Engineering (Các hàm toán học thường dùng trong kỹ thuật),...

- Nửa bên phải là tên các hàm của Categories đó. Chọn 1 hàm thích hợp sau đó vào hộp thoại để chọn đối số (Argument), đối số là địa chỉ các ô. Ta nhập địa chỉ hay tên miền và tùy theo hàm có 1, 2, hay nhiều đối số mà trả lời cho đủ, phía dưới hộp thoại có chỉ dẫn về nội dung của hàm và nội dung đối số, phần Help giúp ta tìm hiểu kỹ hơn kể cả thí dụ của EXCEL.

* Cách khai thác lệnh

- Vào Insert/ Function ta vào hộp thoại của các hàm định sẵn tương tự như vào f_x .

- Vào Data/ Pivot Table: Để tổng hợp dữ liệu theo nhiều tiêu thức khác nhau; giải quyết đồng thời cả phân tổ thống kê kết hợp theo nhiều tiêu thức và tính toán các chỉ tiêu theo các hàm thống kê đơn giản.
- Vào Tools/Data Analysis: Để phân tích dữ liệu như:
 - + So sánh hai hoặc nhiều số bình quân bằng phương pháp phân tích phương sai.
 - + Kiểm định 2 số bình quân theo phân bố T.
 - + Kiểm định 2 tập hợp số liệu xem 2 phương sai của chúng như thế nào ?
 - + Giải bài toán tương quan và hồi quy tuyến tính.
 - + Tính toán ma trận riêng giữa các yếu tố.
 - + Xác định các chỉ tiêu thống kê đơn giản v.v....

2.3 Phần mềm SPSS

a) Giới thiệu chung về SPSS

- SPSS là phần mềm chuyên ngành thống kê khởi đầu vào những năm 1960, trong môi trường DOS.
- SPSS 3.0 có từ 1972, chạy trên môi trường DOS, đòi hỏi bộ nhớ 4 Mb.
- Năm 1992 thì SPSS 5.0 xuất hiện.
- Công ty sở hữu phần mềm (Chicago - Mỹ) đã biên soạn 7 cuốn sách hướng dẫn sử dụng theo các chuyên đề khác nhau (mỗi cuốn dày từ 500 trang đến 1000 trang) cho từng phiên bản SPSS.
- Hiện nay đã xuất hiện trên thị trường SPSS phiên bản 15, giá của mỗi phiên bản phần mềm sử dụng cho cá nhân khoảng 700 USD.
- Phiên bản 6 dùng cho Windows 3.x còn phiên bản 7 trở lên dùng cho Windows 95 - 97, 98, 2000.
- SPSS viết tắt của “Statistical Package for the Social Sciences (Phần mềm thống kê chọn gói cho các môn khoa học xã hội).
- Mặt khác, SPSS cũng là viết tắt của Statistical Products for the Social Services (Các sản phẩm thống kê cho các dịch vụ xã hội).
- SPSS là một phần mềm chọn gói mang tính tổng hợp chung, có đủ khả năng để giải quyết các nhiệm vụ phức tạp.
- SPSS là một phần mềm thống kê toàn diện được thiết kế để thực hiện tất cả các bước trong bất kỳ phân tích thống kê nào từ việc liệt kê dữ liệu, lập bảng biểu và các thống kê mô tả đến các phân tích thống kê phức tạp hơn.
- SPSS được thiết lập rất tốt trên môi trường Windows và rất dễ dàng cho người sử dụng, không phải lập trình để giải quyết những nhiệm vụ phức tạp, làm cho quá trình phân tích dữ liệu trở nên dễ hơn và ngắn gọn hơn.
- SPSS for Windows có rất nhiều nét chung với các phần mềm tiện ích khác dùng cho Windows như EXCEL hay Winword.
- Phiên bản SPSS dùng trong môi trường Windows có một tính năng ưu việt so với SPSS trên DOS là có mục HELP (trợ giúp) tương đối đầy đủ để giúp người sử dụng tự giải quyết những khó khăn vướng mắc (tất nhiên lợi thế này chỉ có người thành thạo tiếng Anh, thành thạo Tin học, thành thạo về chuyên môn mới khai thác hết được).

b) Sự cần thiết ứng dụng SPSS

- So với các phần mềm khác như EXCEL, SPSS là phần mềm đa năng, cho phép tổng hợp, phân tích và xử lý các dữ liệu kinh tế xã hội đầy đủ và tiện lợi.
- SPSS có thể cho ra các bảng tính tần suất của tất cả các biến (định tính và định lượng) trong cơ sở dữ liệu hoặc lập một bảng tương quan giữa hai biến với cách trình bày đẹp.

Ví dụ, trong cơ sở dữ liệu của một cơ quan có thể lập các bảng tổng hợp: Mức lương phân theo độ tuổi, mức lương phân theo trình độ học vấn, đoàn thể phân theo giới tính, .. Chỉ trong vòng vài phút, có thể tạo được hàng trăm biểu bảng tổng hợp mà không phải lập trình như nhiều chương trình khác.

- SPSS cũng thích hợp cho những người không cần biết sâu về máy tính, phần mềm này có thể liên kết các phần mềm khác như EXCEL, STATA, SAS, v.v...

- SPSS thường được dùng để xử lý các cuộc điều tra dân số, các cuộc điều tra xã hội học, điều tra về tình hình kinh tế, thị trường, phân tích các cơ sở dữ liệu. SPSS được dùng nhiều trong khoa học xã hội và thống kê.

c) Một số thuật ngữ trong SPSS

- **Case:** Case là các quan sát hay bản ghi hay trường hợp hoặc chủ thể. Một trường hợp bao gồm các thông tin cho một đơn vị của phép phân tích, chẳng hạn như một người, một con vật, một công việc, hoặc một động cơ phản lực.

- **Variable (biến, tiêu thức):** Variable là các thông tin (đặc điểm) được thu thập cho từng trường hợp. Ví dụ, chiều cao, trọng lượng, lợi nhuận hay doanh số. Mỗi trường hợp được tạo nên bởi các biến. Mỗi biến được đặt tên biến cụ thể (variable name) với tối đa là 8 ký tự. Tuy nhiên, bạn cũng có thể chỉ định nhãn biến (variable label) để mô tả cho tên đầy đủ của biến, chỉ định giá trị biến (value label) cho những biểu hiện riêng biệt của từng biến (ví dụ như Nam = 1, Nữ = 0). Bạn cũng có thể tùy chọn hoặc là nhãn biến hoặc là tên biến trong kết xuất đầu ra của mình.

- **SPSS working data file (tệp hiện hành SPSS):** Các trường hợp tập hợp lại với nhau sẽ tạo nên tệp hiện hành SPSS.

- **Dictionary (thư mục):** Chứa các tệp dữ liệu, cả tệp kết quả.

- **Metadata (siêu dữ liệu):** Các siêu dữ liệu được chứa trong mỗi tệp dữ liệu cụ thể, chẳng hạn như tên các biến, các nhãn, định dạng v.v.. là các siêu dữ liệu.

- **Đuôi mở rộng:** *.sav; *.sps; *.spo

- **Measurement (thang đo):** Các biểu hiện (giá trị) của các biến được xác định bằng các thang đo khác nhau tùy theo tính chất của việc đo lường.

Có 4 loại thang đo thường gặp:

+ **Nominal (thang đo định danh):** Là đánh số hoặc gán chuỗi dạng ngắn cho các biểu hiện của một biến được gọi là biến định danh (nominal variable).

Các trị số của biến định danh chỉ biểu hiện các nhóm không có thứ bậc hơn kém (unordered categories). Nếu biến định danh được đo bằng các con số thì giữa các con số ở đây không có quan hệ hơn kém, do vậy mọi phép tính đại số với chúng đều vô nghĩa. Ví dụ, hai biểu hiện của biến giới tính có thể được biểu hiện như sau: 1 = nam và 0 = nữ. Thang đo định danh chủ yếu để đếm tần số của biểu hiện của biến nghiên cứu. Trong nhiều phép phân tích của SPSS, các biến định danh thường được gọi là các biến lập nhóm không có thứ bậc (unordered categorical variable).

+ **Ordinal (thang đo thứ bậc):** Là thang đo định danh nhưng các trị số của biến lại có quan hệ thứ bậc hơn kém.

Ví dụ mức thu nhập có thể xác định bằng 3 biểu hiện: 1 = thấp, 2 = trung bình, 3 = cao. Sở thích nhạc cổ điển có thể xác định bằng 5 biểu hiện: 1 = rất thích, 2 = thích, 3 = bình thường, 4 = ghét và 5 = rất ghét. Con số có trị số cao hơn không có nghĩa là luôn ở bậc cao hơn và ngược lại mà chỉ do sự quy định. Thang đo này dùng để tính toán đặc trưng chung của tổng thể một cách tương đối, ví dụ như ở sở thích về một thể loại âm nhạc của công chúng.

Các biến được đo đặc bằng thang đo thứ bậc gọi là các biến định danh có thứ bậc (ordinal variable). Trong nhiều phép phân tích của SPSS, các biến định danh có thứ bậc thường được gọi là các biến lập nhóm có thứ bậc (ordered categorical variable).

+ **Interval (thang đo khoảng):** Là thang đo thứ bậc có khoảng cách đều nhau. Ví dụ như chiều cao (cm), cân nặng (kg), tuổi thọ của con người (năm), doanh số (triệu đồng). Các phép tính số học như cộng, trừ, bình quân, hoặc phương sai là có ý nghĩa với các biến được đo đạc bằng thang đo khoảng.

Loại thang đo khoảng bao hàm cả hai ý nghĩa về thứ tự của hạng loại và sự khác biệt đo đạc giữa chúng. Nó hàm ý sự hiện diện một đơn vị chung của việc đo đạc tiêu chuẩn mà qua đó khoảng cách giữa các giá trị có thể được lượng hoá. Giá trị số liệu nằm trong khoảng số thực (real interval) và có thể biến thiên từ âm không xác định đến dương không xác định. Sự khác biệt giữa hai số liệu khoảng là có ý nghĩa nhưng tỷ lệ của chúng thì không có ý nghĩa. Thí dụ, chỉ số IQ khi nói rằng người này thông minh hơn người kia là 1,2 lần thì không có ý nghĩa.

+ **Ratio (thang tỉ lệ):** Là thang đo khoảng, phản ánh sự khác biệt và tỷ lệ của 2 giá trị có ý nghĩa. Thí dụ tiền lương, chiều cao, trọng lượng, tỷ trọng lao động trực tiếp, v.v...

Trong SPSS các thang đo khoảng cách, tỷ lệ được gọi chung là kiểu Scale. Ngoài ra, căn cứ vào thang đo của chúng các biến có thể được phân loại thành biến định tính và biến định lượng.

- **Attribute variable (biến định tính):** Là biến lập nhóm (categorical variable) vì trị số của nó được xác định bằng thang đo định danh hoặc thang đo thứ bậc dưới dạng mã số hoặc chuỗi ngắn.

Cần nhớ rằng trong các hộp thoại của SPSS, các trị số của biến định tính dạng chuỗi luôn luôn được tham chiếu trong các dấu ngoặc đơn hoặc dấu ngoặc kép.

- **Quantitative variable (biến định lượng):** Luôn luôn được đo đạc theo thang đo khoảng nên trị số của nó luôn được biểu hiện dưới dạng số.

- **Menu bar (thanh trình đơn chính):** Chúng là các trình đơn kéo xuống (pull - down menu) và mỗi trình đơn phụ có thể có nhiều lỗi vào.

- **Tools bar (thanh công cụ):** Cung cấp truy cập nhanh lựa chọn trình đơn.

- **Status bar (thanh tình trạng):** Nằm ở dưới đáy của từng cửa sổ cho biết trạng thái hiện hành của bộ xử lý của SPSS (SPSS processor).

- **Data (dữ liệu):** SPSS cho phép chúng ta đọc rất nhiều loại tệp dữ liệu hoặc nhập dữ liệu trực tiếp vào trình Hiệu chỉnh dữ liệu (SPSS Data Editor).

Bất kể cấu trúc của tệp dữ liệu gốc của bạn như thế nào, dữ liệu trong trình Hiệu chỉnh dữ liệu đều được trình bày theo cấu trúc hình chữ nhật - định dạng của SPSS và hầu hết các hệ thống phân tích dữ liệu - các hàng cho các trường hợp (case) và các cột cho các biến (variable).

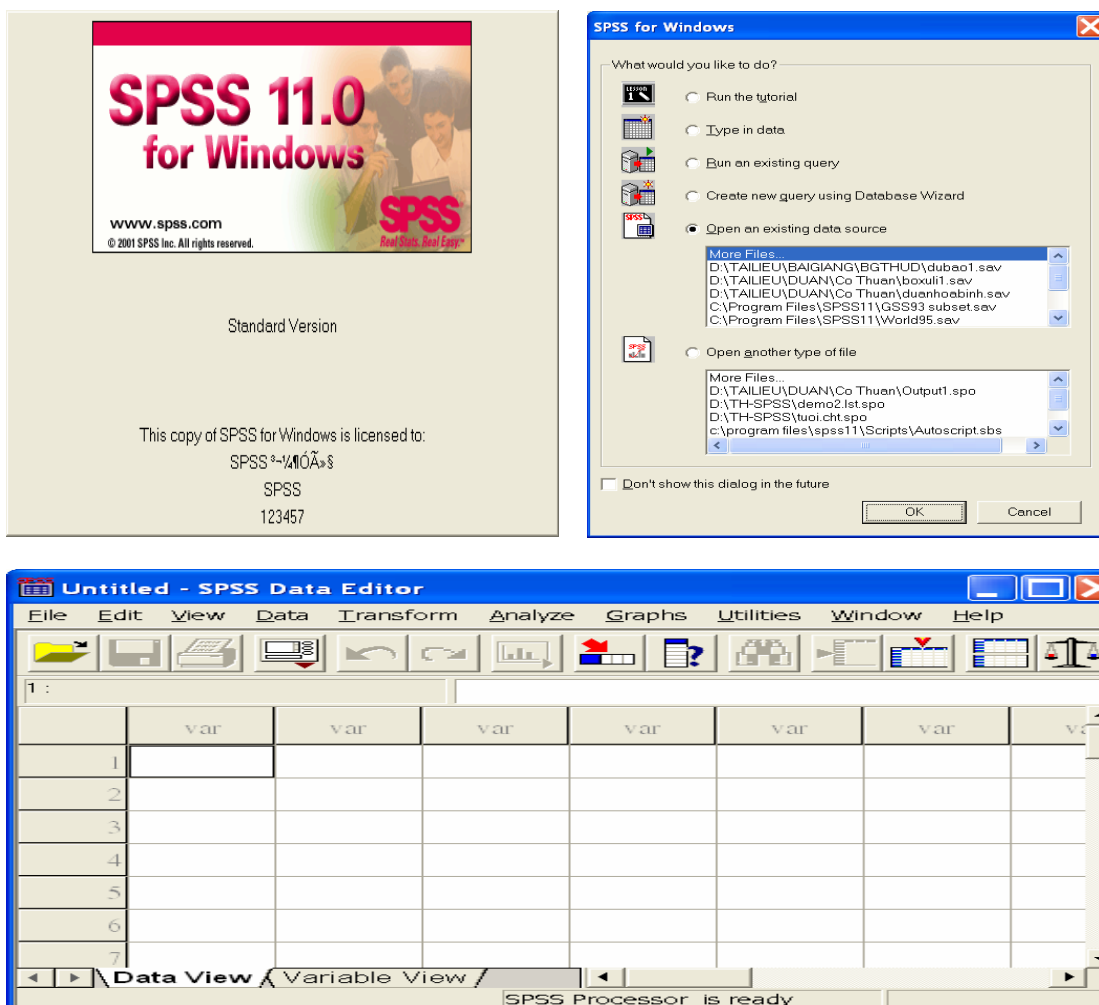
- **Dialogue box (các hộp thoại):** Các hộp thoại có đặc tính thông dụng của Windows như các nút ấn (pushbuttons), các hộp danh sách (list boxes), hộp lựa chọn (option boxes) và các hộp kiểm tra (check boxes).

Các hộp thoại được thiết lập là duy trì liên tục qua từng lần làm việc của SPSS, có nghĩa là các lựa chọn đã được thực hiện sẽ được duy trì khi bạn mở cùng hộp thoại ở lần tiếp theo.

d) Một số thao tác cơ bản

* Khởi động

Start/Program/SPSS for windows hoặc kích vào biểu tượng tương ứng ta nhận được màn hình soạn thảo số liệu của SPSS như sau:



SPSS for Windows cung cấp một hệ thống quản lý dữ liệu và phân tích thống kê trong một môi trường đồ họa, sử dụng các trình đơn mô tả {menu} và các hộp thoại {dialogue box} đơn giản để thực hiện hầu hết các công việc cho bạn. Phần lớn các nhiệm vụ có thể được hoàn thành chỉ bằng cách rê và nhấp chuột.

Bên cạnh giao diện rê-nhấp chuột để phân tích thống kê, SPSS for Windows cung cấp:

Data Editor {Cửa sổ Hiệu đính dữ liệu}. Một hệ thống dạng bảng tính {worksheet} uyển chuyển để định nghĩa, nhập, hiệu đính, và thể hiện dữ liệu.

Viewer {Cửa sổ Viewer}. Cửa sổ Viewer cho phép dễ dàng duyệt các kết quả của bạn, thể hiện và che giấu có thể chọn lọc các kết xuất {output}, thay đổi trật tự của các kết quả, và di chuyển các bảng và đồ thị giữa SPSS for Windows và các trình ứng dụng khác.

Multidimension pivot table {Bảng trụ đa chiều}. Các kết quả của bạn sẽ sinh động với các bảng trụ đa chiều. Khám phá các bảng của bạn bằng cách bố trí lại các hàng, các cột, và các trang/lớp {layer}. Bộc lộ các phát hiện quan trọng có thể bị mất trong các báo cáo tiêu chuẩn. So sánh các nhóm dễ dàng bằng cách chia tách bảng của bạn sao cho mỗi lần chỉ có một nhóm được thể hiện.

High-revolution graphics {Đồ thị có độ phân giải/độ nét cao}. Các biểu đồ hình tròn, đồ thị cột, biểu đồ tần suất, đồ thị phân tán có độ phân giải cao, màu sắc sống động, các đồ thị ba chiều, và hơn thế nữa được bao gồm như là các tính năng chuẩn trong SPSS.

Database access {Truy cập dữ liệu}. Truy cập dữ liệu từ các cơ sở dữ liệu bằng cách sử dụng trình chỉ dẫn Database Wizard thay vì các truy vấn SQL phức tạp.

Data transformation {Biến đổi dữ liệu}. Tính năng biến đổi dữ liệu giúp bạn có được dữ liệu sẵn sàng cho các bước phân tích. Bạn có thể dễ dàng nhóm, bổ sung, tổng hợp, trộn, chia và chuyển đổi file, và hơn thế nữa.

Các cửa sổ trong SPSS

Có một số loại cửa sổ khác nhau trong SPSS:

Data Editor. Cửa sổ này thể hiện nội dung của file dữ liệu. Bạn có thể lập một file dữ liệu mới hoặc hiệu chỉnh thay đổi một file đã có sẵn với cửa sổ Data Editor. Cửa sổ Data Editor tự động mở ra khi bạn kích hoạt/khởi động SPSS. Bạn chỉ có thể một file dữ liệu tại một thời điểm mà thôi (không thể mở hơn một file dữ liệu vào cùng một thời điểm).

Viewer. Mọi kết quả thống kê, bảng, biểu đồ được thể hiện trong cửa sổ Viewer. Bạn có thể hiệu đính kết xuất và lưu nó để sử dụng sau này. Một cửa sổ Viewer tự động mở ra khi bạn chạy một thủ tục đầu tiên tạo nên kết xuất.

Draft Viewer. Bạn có thể trình bày kết xuất như là các văn bản bình thường (thay vì các bảng trụ) trong cửa sổ Draft Viewer.

Pivot Table Editor. Kết xuất được trình bày trong các bảng trụ có thể được chỉnh sửa bằng nhiều cách với cửa sổ Pivot Table Editor. Bạn có thể hiệu đính đoạn văn bản, chuyển đổi dữ liệu giữa hàng và cột, bổ sung màu, tạo các bảng đa chiều và ẩn hoặc hiển thị một cách có chọn lọc các kết quả.

Chart Editor. Bạn có thể chỉnh sửa các đồ thị chất lượng cao trong các cửa sổ chart editor. Bạn có thể thay đổi màu, chọn loại phong hoặc cỡ chữ, chuyển đổi trục tung với trục hoành, xoay các đồ thị ba chiều, và thậm chí thay cả loại đồ thị.

Text Output Editor. Các kết xuất dạng văn bản không được thể hiện trong các bảng trụ có thể được chỉnh sửa với cửa sổ Text Output Editor. Bạn có thể hiệu đính kết xuất và thay các thuộc tính của phong chữ (dạng, loại, màu, cỡ).

Syntax Editor. Bạn có thể dán các lựa chọn trong các hộp thoại vào một cửa sổ syntax, nơi mà các lựa chọn của bạn xuất hiện dưới dạng các cú pháp lệnh. Bạn có thể hiệu đính các cú pháp lệnh để tận dụng các đặc tính đặc biệt của SPSS không có sẵn trong các hộp thoại. Bạn cũng có thể lưu các mã lệnh này trong một file để sử dụng cho những công việc tiếp theo của SPSS.

Script Editor. Kỹ thuật tự động OLE cho phép bạn tùy biến và tự động hoá nhiều nhiệm vụ trong SPSS. Sử dụng cửa sổ Script Editor để lập và hiệu đính các trình nhỏ cơ bản.

Thanh menu {Menu}

Rất nhiều nhiệm vụ bạn muốn tiến hành với SPSS bắt đầu với việc lựa chọn các menu {trình đơn}. Từng cửa sổ trong SPSS có các menu riêng của nó với các lựa chọn menu thích hợp cho loại cửa sổ đó.

Hai menu Analysis và Graphs là có sẵn đối với mọi loại cửa sổ, làm cho việc tạo các kết xuất mới rất nhanh chóng mà không phải chuyển đổi giữa các cửa sổ.

Thanh công cụ {Toolbars}

Từng cửa sổ SPSS có các thanh công cụ riêng của nó cho phép truy cập nhanh đến các nhiệm vụ thông dụng. Có một số cửa sổ có hơn một thanh công cụ.



Thanh tình trạng {Status Bar}

Thanh tình trạng {status bar} nằm ở đáy của từng cửa sổ SPSS cung cấp các thông tin dưới đây:

Command status {Tình trạng lệnh}. Đối với từng lệnh hoặc thủ tục mà bạn chạy, một số đếm các đối tượng/trường hợp {case} chỉ ra số lượng các đối tượng được xử lý. Đối với các thủ tục đòi hỏi phải xử lý lặp, số lần lặp được thể hiện.

Filter status {Tình trạng lọc}. Nếu bạn chọn một mẫu ngẫu nhiên hoặc một tập hợp phụ các đối tượng để phân tích, thông tin Filter on chỉ ra rằng một vài nhóm đối tượng nào đó đang được lọc và không phải mọi đối tượng trong tệp tin dữ liệu được đưa vào phân tích.

Weight status {Tình trạng gia quyền}. Thông tin Weight on chỉ ra rằng một biến gia quyền đang được sử dụng để gia quyền các đối tượng cho phân tích.

Split status {Tình trạng chia tách}. Thông tin Split on chỉ ra rằng file dữ liệu đang được chia tách thành một số nhóm để phân tích, được dựa vào các trị số của một hoặc một số biến lập nhóm/phân tổ.

Hộp thoại {Dialogue box}

Hầu hết các lựa chọn menu mở ra các hộp thoại. Bạn sử dụng hộp thoại để lựa chọn các biến và các tùy chọn cho phân tích

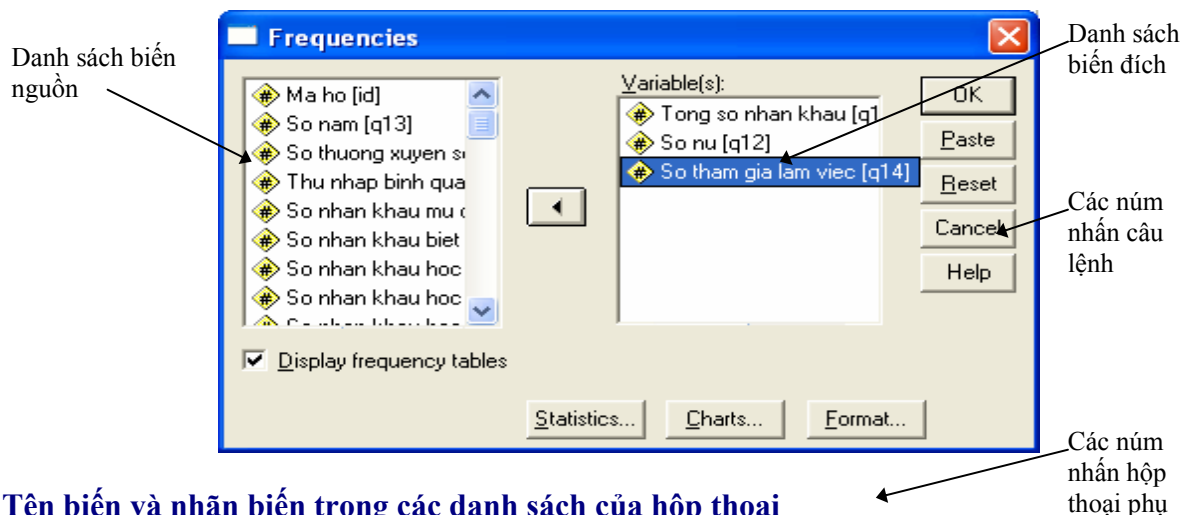
Từng hộp thoại chính cho các thủ tục thống kê và đồ thị có một số các bộ phận cơ bản.

Danh sách biến nguồn. Một danh sách các biến trong file dữ liệu làm việc. Chỉ có các loại biến được phép bởi các thủ tục được chọn mới được thể hiện trong danh sách nguồn. Việc dùng các biến chuỗi dạng ngắn hay dài bị hạn chế bởi rất nhiều thủ tục.

Danh sách (hoặc các danh sách) biến đích. Một hoặc một vài danh sách thể hiện các biến bạn vừa chọn cho phân tích, chẳng hạn như danh sách biến độc lập và phụ thuộc.

Nút ấn điều khiển {Command pushbutton}. Các nút chỉ dẫn chương trình thực hiện một tác vụ, chẳng hạn như chạy một thủ tục, thể hiện phần thông tin Trợ giúp, hoặc mở ra một hộp thoại con để tiến hành các lựa chọn cụ thể bổ sung.

Để có được thông tin về các nút điều khiển trong một hộp thoại, nhấp chuột phải lên nút đó.



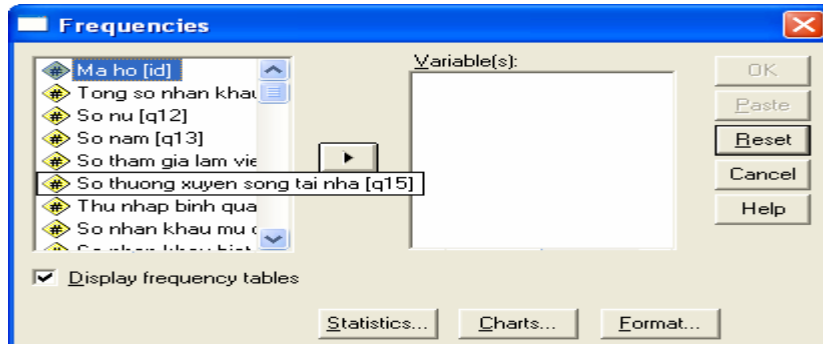
Tên biến và nhãn biến trong các danh sách của hộp thoại

Bạn có thể thể hiện hoặc là tên biến hoặc là nhãn biến trong danh sách của hộp thoại. Do tên biến bị hạn chế bởi 8 ký tự, nhãn biến thường cung cấp nhiều thông tin mô tả biến hơn.

- Để điều khiển sự thể hiện tên biến hay nhãn biến trong danh sách của hộp thoại, trong Options trong menu Edit ở bất kỳ loại cửa sổ nào của SPSS.

- Để định nghĩa hoặc chỉnh sửa nhãn biến, hãy nhấp đúp tên biến trong cửa sổ Data Editor và sau đó nhấp Labels.

- Đối với dữ liệu nhập từ các nguồn cơ sở dữ liệu, tên các trường được sử dụng làm nhãn biến.
- Đối với nhãn biến quá dài, chỉ con trỏ lên nhãn trong danh sách để xem toàn bộ nhãn biến đó.
- Nếu không có nhãn biến nào được xác định thì tên biến sẽ được thể hiện.



Các nút trong hộp thoại

Có 5 nút nhấn tiêu chuẩn trong hầu hết các hộp thoại:

- **OK.** Chạy thủ tục. Sau khi bạn chọn các biến nghiên cứu và chọn bất kỳ các tùy chọn bổ sung nào, nhấp OK để chạy thủ tục. Điều này cũng đồng thời đóng hộp thoại lại.
- **Paste.** Tạo cú pháp câu lệnh từ các lựa chọn trong hộp thoại và dán cú pháp vào một cửa sổ cú pháp. Sau đó bạn có tùy biến các câu lệnh với các đặc tính bổ sung không có sẵn trong hộp thoại.
- **Reset.** Bỏ chọn bất kỳ biến nào trong danh sách các biến được chọn và thiết lập mặc định cho mọi tùy chọn trong hộp thoại và bất kỳ hộp thoại phụ nào.
- **Cancel.** Xóa bỏ bất kỳ thay đổi nào trong thiết lập hộp thoại kể từ lần cuối nó được mở ra và đóng hộp thoại lại. Trong mỗi lần làm việc với SPSS các thiết lập trong hộp thoại là luôn tồn tại cho đến khi bạn thoát khỏi SPSS. Một hộp thoại duy trì mọi thiết lập mà bạn chọn cho đến khi bạn thiết lập lại. .
- **Help.** Nút này cho bạn cửa sổ trợ giúp dạng chuẩn của hãng Microsoft bao gồm các thông tin về hộp thoại hiện tại. Bạn cũng có thể nhận được các trợ giúp trong các nút điều khiển riêng trong từng hộp thoại bằng cách nhấp chuột phải lên nó.

Hộp thoại phụ

Do hầu hết các thủ tục đều cung cấp một sự uyển chuyển lớn, không phải mọi lựa chọn đều có thể được bao hàm chỉ trong một hộp thoại. Hộp thoại chính bao gồm các thông tin tối thiểu đòi hỏi để chạy một thủ tục. Các thiết lập bổ sung được thực hiện trong các hộp thoại phụ.

Trong hộp thoại chính, nút nhấn với ba dấu chấm (...) đằng sau tên của nó chỉ ra rằng một hộp thoại phụ sẽ được xuất hiện nếu bạn nhấp chuột vào nó.

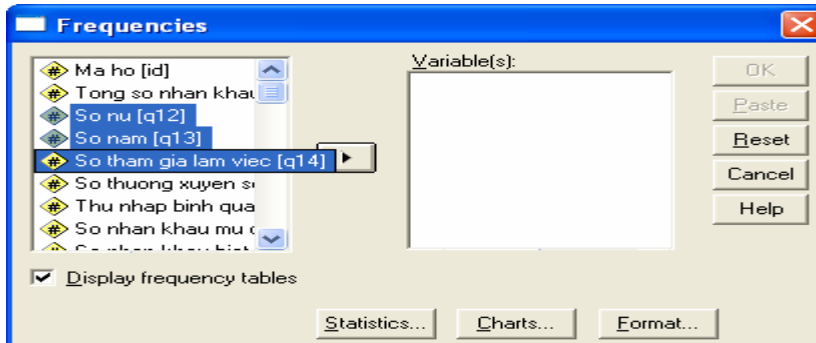
Lựa chọn biến

Để lựa chọn một biến, bạn chỉ đơn giản nhấp chuột vào nó trong danh sách các biến nguồn và nhấp nút mũi tên phải nằm bên cạnh danh sách các biến nguồn. Nếu chỉ có một danh sách các biến nguồn, bạn có thể nhấp đúp các biến đơn để chuyển chúng từ danh sách nguồn sang danh sách tới.

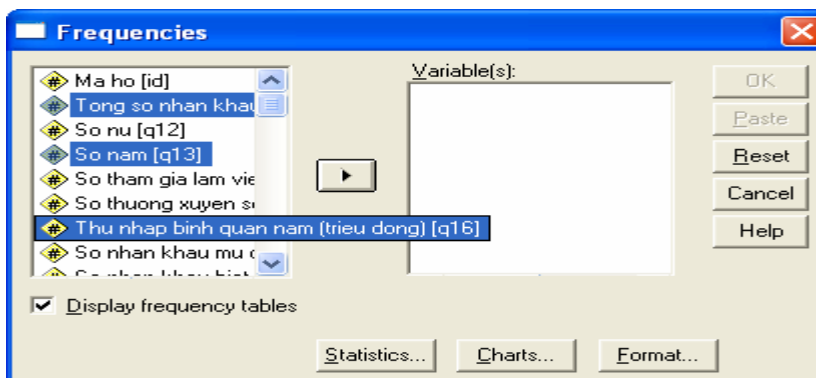
Bạn có thể chọn nhiều biến một lúc:

- Để chọn nhiều biến nằm kế nhau liên tục trong danh sách các biến nguồn, nhấp vào biến đầu tiên và giữ phím Shift và nhấp vào biến cuối cùng.

- Để chọn các biến không nằm kề nhau liên tục (nằm cách quãng) trong danh sách các biến nguồn, hãy sử dụng phương pháp nhấp+Ctrl. Chọn biến đầu tiên, sau đó giữ phím Ctrl và nhấp biến tiếp theo, và cứ thế tiếp tục cho đến biến cuối cùng.
- Để chọn mọi biến trong danh sách, nhấn Ctrl+A
- + Lựa chọn nhiều biến với kỹ thuật Shift cùng với nhấp chuột

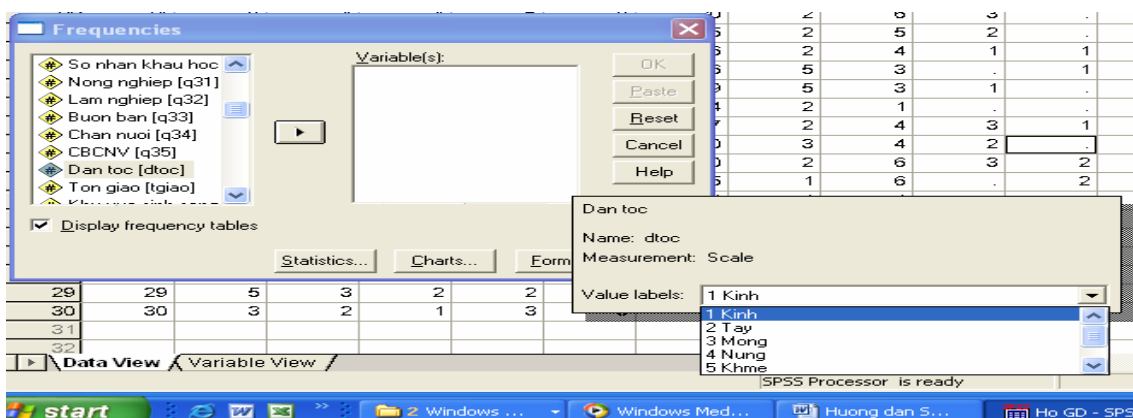


- + Chọn nhiều biến kế tiếp nhau với kỹ thuật Ctrl cùng với nhấp chuột



Để có được thông tin về một biến trong một danh sách trong một hộp thoại

- ▶ Nhấp chuột trái lên một biến trong một danh sách để chọn nó
- ▶ Nhấp chuột phải bất kể nơi nào trong danh sách
- ▶ Chọn Variable Information trong menu pop-up
- + Xem thông tin về biến dùng phím chuột phải

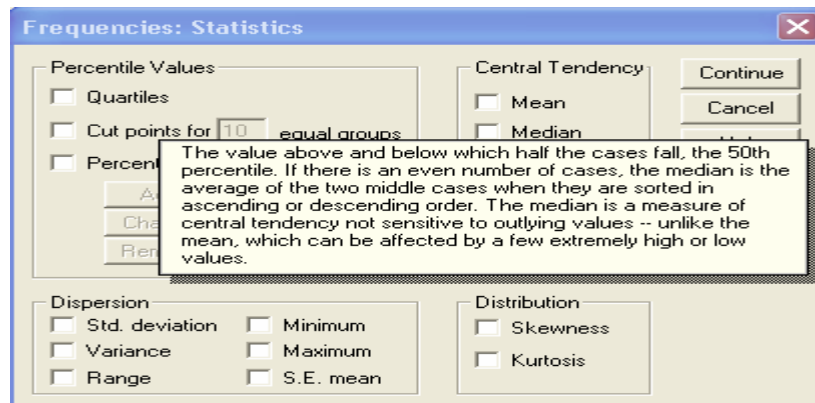


Để nhận được thông tin về nút điều khiển trong hộp thoại

- ▶ Nhấp chuột trái lên nút bạn muốn biết
- ▶ Chọn What's This? Trong menu pop-up.

Một cửa sổ pop-up thể hiện thông tin về nút điều khiển.

+ Trợ giúp dạng “What’s This?” pop-up bằng cách nhấp phím phải chuột



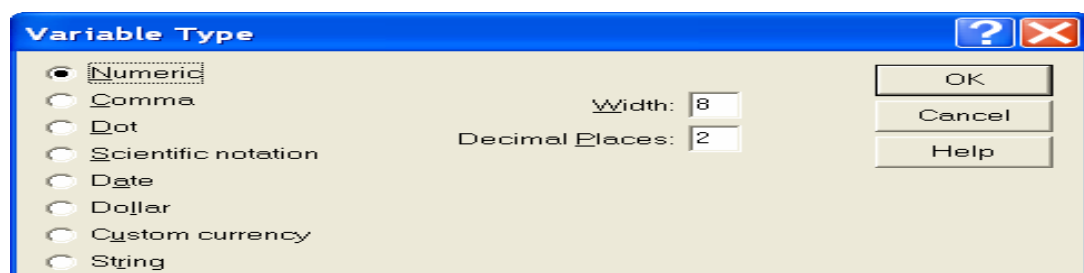
* **Thoát khỏi SPSS:** File/ Exit SPSS hoặc ấn đồng thời Alt + F4

* **Nhập số liệu trực tiếp:** Chọn Variable View từ trang màn hình của SPSS

+ **Khai báo tên biến (Name)**

Tên biến khai báo theo quy định thông thường (không quá 8 kí tự, không bắt đầu bằng số). Số biến tối đa chấp nhận được của một tệp số liệu là 10.000 biến.

+ **Khai báo kiểu biến (Type)**

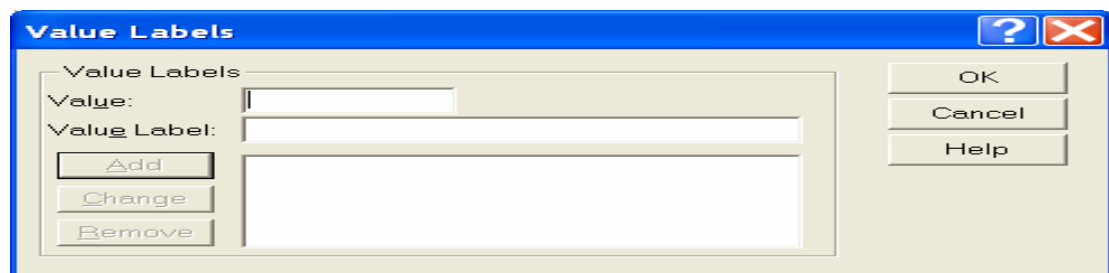


+ **Khai báo độ rộng của biến (Width)**

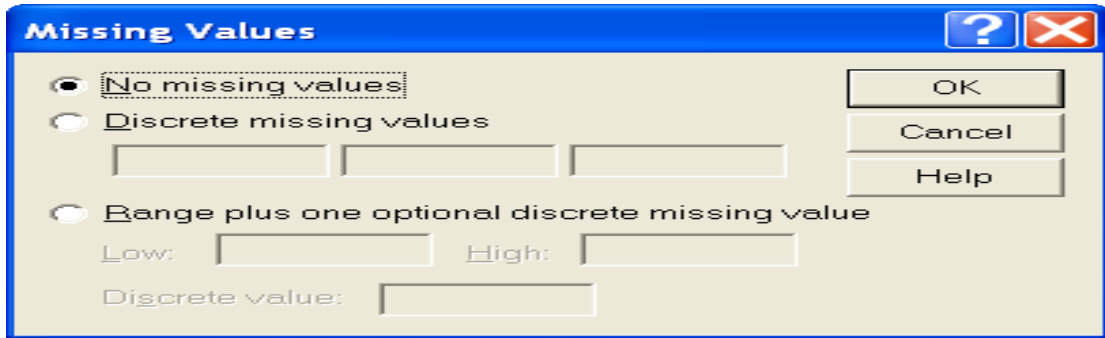
+ **Khai số chữ số đằng sau dấu phẩy (số chữ số thập phân) (Decimals)**

+ **Khai báo nhãn biến (Label)**

+ **Khai báo giá trị của nhãn (mã hoá) (Values)**



+ **Khai báo giá trị thiếu của biến (Missing)**



Tại sao phải khai báo giá trị thiếu của biến ?

Một số biến có thể không có giá trị trong một vài bản ghi, vì vậy cần khai báo các giá trị cho các địa chỉ thiếu này, nếu biến này là một biến định lượng thì sau này ta có thể sử dụng một trong các thuật toán để bổ sung các giá trị thiếu, tuy nhiên việc chọn giá trị thiếu nói chung là cần thiết ngay khi vào sổ liệu. Giá trị thiếu mặc định của SPSS là một dấu chấm (.), có thể đặt giá trị thiếu cho từng biến khác nhau nhưng rất chú ý đến khả năng kết hợp sau này của các biến, vì vậy cần thận trọng trong việc chọn các giá trị này.

+ Khai báo độ rộng của cột (columns)

+ Căn lề cột (Align): Left, Right, Center

+ Chọn thang đo (Measure): Scale, Ordinal, Nominal

+ Sau đó chuyển sang Data View để nhập số liệu trực tiếp trên bàn phím

* Copy số liệu từ một trang bảng tính khác

Có thể một số người chưa làm quen với SPSS tỏ ra không tự tin khi nhập số liệu trực tiếp với hàng loạt khai báo như trên. Chúng ta dễ dàng copy số liệu từ một trang tính như bảng tính điện tử Excel để chuyển các số liệu này vào SPSS chỉ cần mở cả SPSS và Excel chọn từng cột cần copy từ Excel sau đó copy vào trang sổ liệu của SPSS rồi tiến hành khai báo biến sau. Cách làm này không phổ biến nhưng trường hợp đặc biệt thì được xử lý nhằm đỡ tốn công sức nhập dữ liệu trực tiếp trong SPSS.

* Đọc số liệu từ một tệp: File/Open

Việc đọc số liệu từ một tệp được ghi bởi SPSS hoặc các phần mềm khác vào SPSS là rất đơn giản. Ta có thể đọc cả tên biến kèm theo hoặc chỉ đọc số liệu và sau đó khai báo biến theo sở thích. SPSS có thể đọc được các tệp của Lotus, Excel, Fox, Dbase, Quattro, Syllk và các tệp ghi bởi mã ASCII.

Để đọc một tệp như vậy, ta chọn **File/ Open** và chọn tệp cần đọc với phần mở rộng tương ứng. Một trong 2 hệ cơ sở dữ liệu lớn là Fox và Dbase có thể đọc hết sức thuận tiện vào SPSS. Thông thường việc đọc dữ liệu từ các phần mềm khác vào SPSS đòi hỏi phải khai báo lại biến trước khi xử lý, tuy nhiên sau này ta sẽ xem xét việc chuyển các dữ liệu đó một cách thuận tiện hơn nhờ thủ tục **Transpose**. Riêng đối với những tệp số liệu lớn trong đó mỗi bản ghi được ghi trên nhiều dòng (mỗi dòng tối đa 256 ký tự) thì việc đọc tệp cần thực hiện nhờ thủ tục của SPSS với một vài khai báo bổ sung, chúng ta có thể xem xét ở phần sau (xem thủ tục đọc tệp indiv.dat). Chẳng hạn, ta đọc tệp vay.sav trong thư mục SPSSWin nhờ các chỉ định: File/ Open/ Solieu/ vay.sav

* Thêm, bớt một biến: Data/Insert variable hoặc chọn biến/Edit/Cut

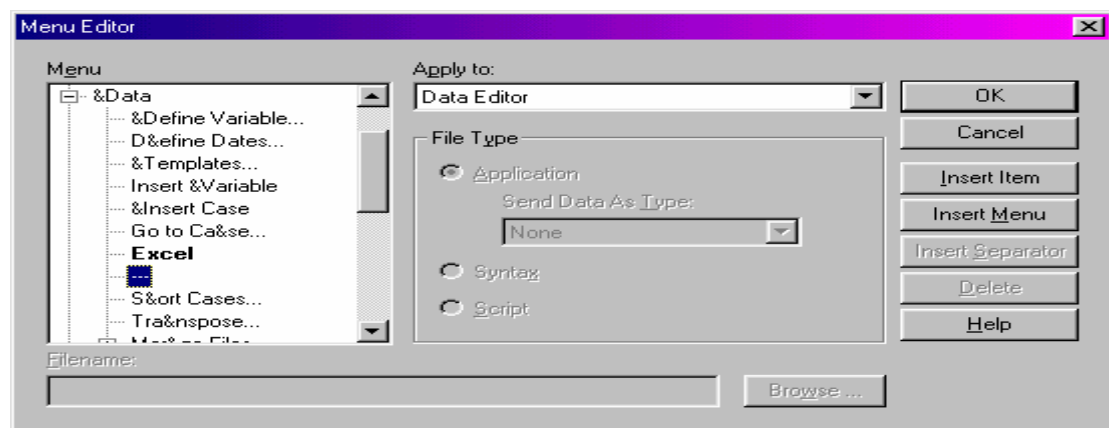
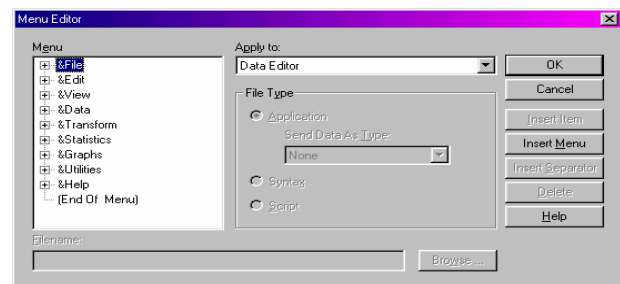
* Ghi số liệu: File/ Save As

Sau khi nhập số liệu từ bàn phím hoặc đọc số liệu từ một tệp được ghi bởi các trình ứng dụng, ta có thể ghi số liệu này thành một tệp của SPSS với phần mở rộng ngầm định là **sav**, thành tệp có phần mở rộng **W***, **XLS**, **XLK**, **DBF**, **DAT** tương ứng với mục đích sử dụng trong các trình ứng dụng khác.

* Chuyển đổi trực tiếp với Excel, Fox, Foxpro, ..

Ngoài cách ghi số liệu đã nói ở trên trong nhiều trường hợp người ta cần một vài xử lý trước trong Excel, Fox, sau đó mới ghi lại thành tệp kiểu XLS hoặc DBF. Có thể chuyển đổi trực tiếp tệp số liệu trong màn hình soạn thảo số liệu đến các ứng dụng khác, đặc biệt là các bảng tính như Lotus, Excel hay các hệ cơ sở dữ liệu như Fox, Foxpro, dBase, .. Để làm được công việc này trước hết cần bổ sung vào thực đơn sẵn có của màn hình soạn thảo số liệu các biểu tượng và đặt đường dẫn cho chúng. Tuy nhiên, khác với việc ghi một tệp số liệu để dùng trong các phần mềm đó, việc chuyển đổi có thể gặp những hạn chế nhất định vì lý do các ứng dụng này không cho phép mô tả số liệu trên một dòng quá dài. Trong trường hợp đó cần cắt tệp theo cách copy trực tiếp thành các tệp cần thiết. Riêng với EXCEL việc làm này khá tiện dụng vì ta có thể mở cùng lúc cả SPSS và EXCEL, hơn nữa bảng tính EXCEL có cùng dạng với màn hình soạn thảo số liệu của SPSS. Sau đây là thủ tục bổ sung biểu tượng EXCEL vào thực đơn Data của SPSS, các biểu tượng khác hoàn toàn tương tự:

- Chọn Utilities/ Menu Editor ta có bảng liệt kê thực đơn Menu Editor
- Chọn mục Data trong liệt kê bên trái
- Chọn vị trí trống trên cây thực đơn và chọn Insert Item để ghi Excel
- Chọn kiểu tệp tương ứng là XLS
- Đặt đường dẫn đến tệp EXCEL.exe chẳng hạn C:\Office\Excel\Excel.exe



(dòng in đậm được chèn thêm nhờ cách làm trên)

* Xem thông tin về tệp: Utilities/ File Info

* Sao chép, cắt dán, xóa số liệu, khôi phục lại: Tương tự như EXCEL

Chú ý: Nguồn là số liệu chứ không phải là công thức, hàm, ... Chúng ta có thể sao chép vùng dữ liệu của SPSS để dán cho Word, Excel,...

* Đối với chèn dòng: Chọn dòng cần chèn/Data/ Insert Case

* **Copy 1 tệp:** Utilities/ File info/ Chọn tệp ở khung phải/ Edit/ Copy/ Chuyển đến phần mềm cần dán (Word, Excel, ...)/ chọn nơi cần dán/ Edit/ Past.

* **Ghi tệp số liệu:** File/ Save As/ ổ đĩa, thư mục, dạng tệp, tên tệp/ Save (đối với SPSS 7.5 trở lên cho phép ghi tên tệp dài được)

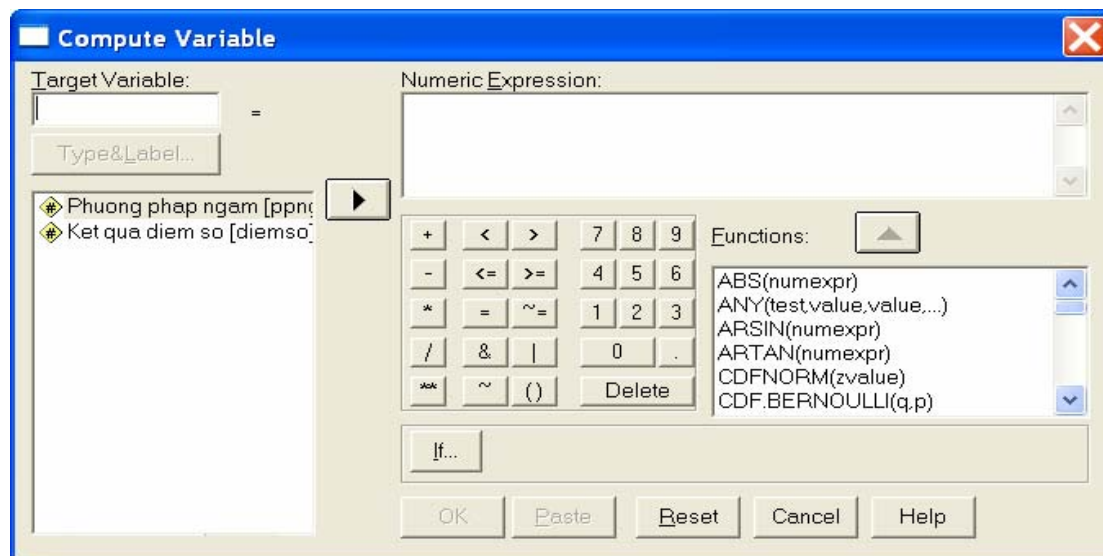
* Xem thông tin của biến: Utilities/ Variables

Chúng ta có thể xem tất cả các thông tin về một biến nào đó từ một bảng liệt kê danh sách biến của tệp số liệu hiện thời.

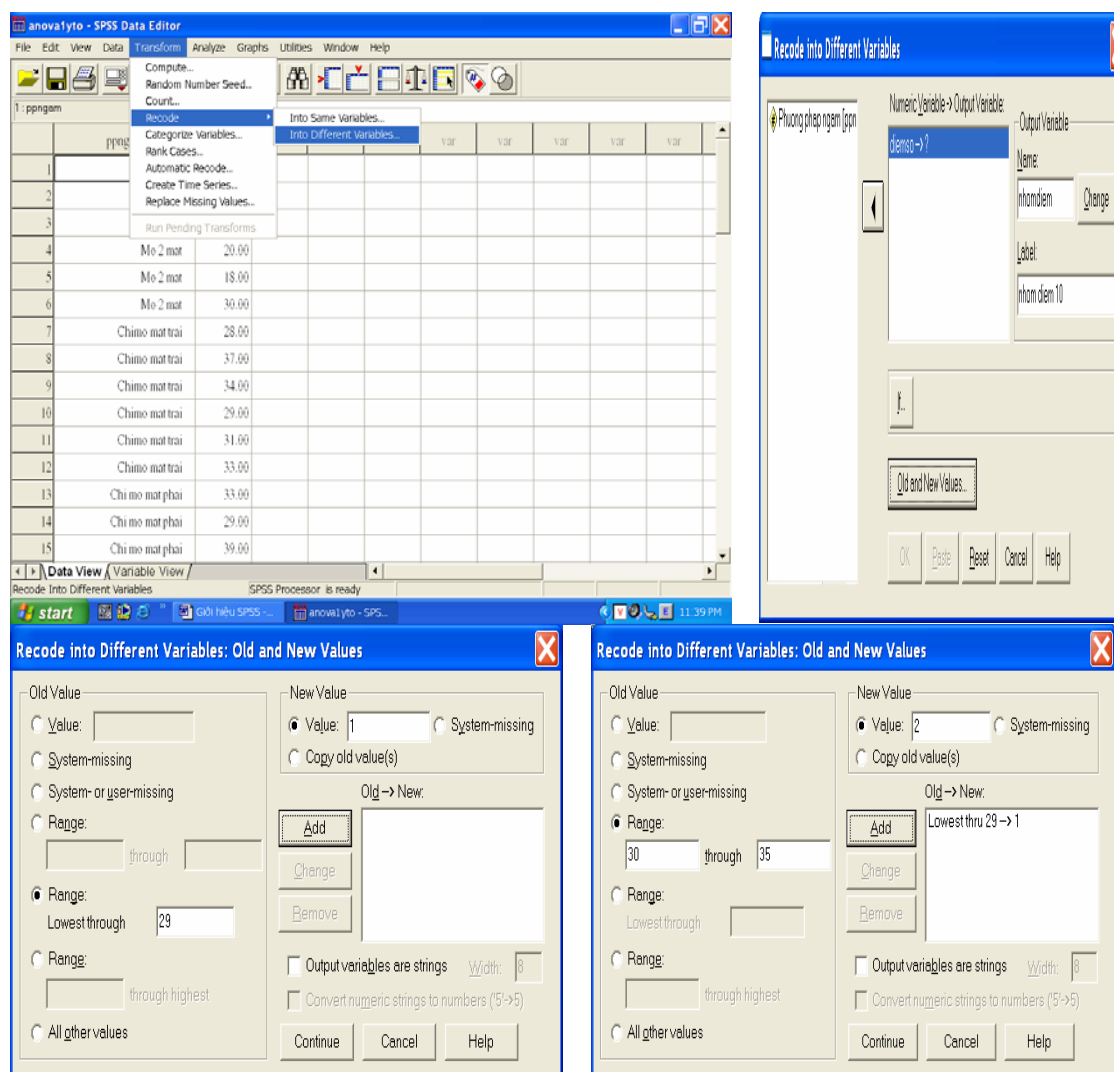
* Sắp xếp số liệu: Data/ Sort case.

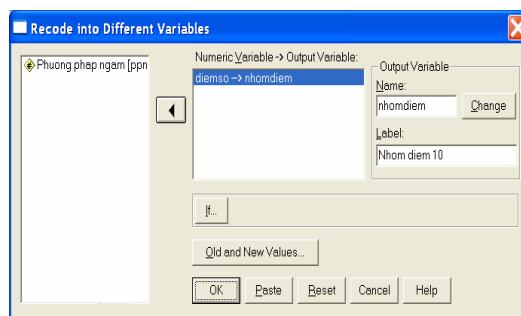
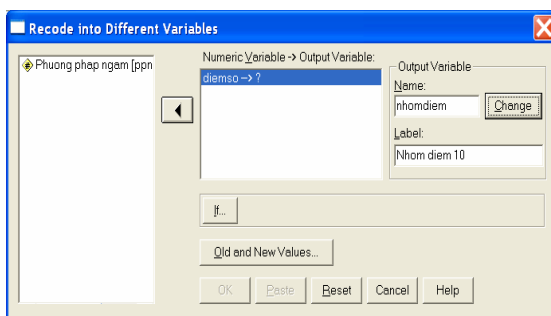
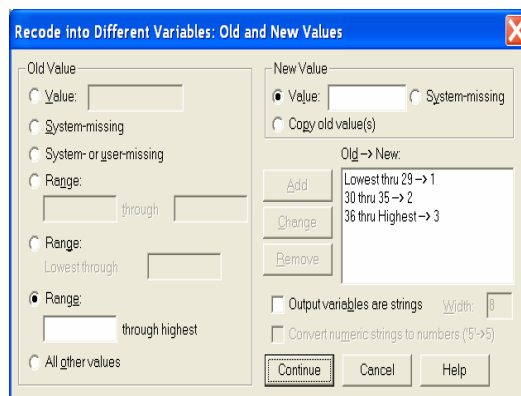
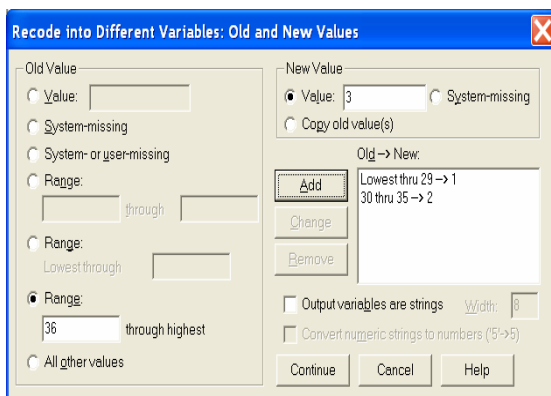
Cho phép sắp xếp số liệu trong tệp theo thứ tự tăng hay giảm của các giá trị biến, số biến cho phép dùng làm khoá để sắp xếp có thể là không hạn chế.

* **Đặt biến: Transform/ Compute.**



* **Nhóm và đặt biến phân nhóm (Mã hoá lại biến): Transform/ Recode**





anova1yto - SPSS Data Editor

	ppngam	diemso	nhomdiem	var	var	var	var	var	var	var
1	Mo 2 mat	22.00	1.00							
2	Mo 2 mat	27.00	1.00							
3	Mo 2 mat	29.00	1.00							
4	Mo 2 mat	20.00	1.00							
5	Mo 2 mat	18.00	1.00							
6	Mo 2 mat	30.00	2.00							
7	Chimo mat trai	28.00	1.00							
8	Chimo mat trai	37.00	3.00							
9	Chimo mat trai	34.00	2.00							
10	Chimo mat trai	29.00	1.00							
11	Chimo mat trai	31.00	2.00							
12	Chimo mat trai	33.00	2.00							
13	Chi mo mat phai	33.00	2.00							
14	Chi mo mat phai	29.00	1.00							
15	Chi mo mat phai	39.00	3.00							

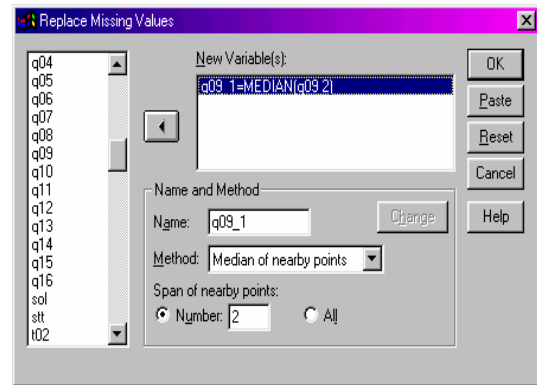
SPSS Processor is ready

* Tạo giá trị cho các giá trị thiếu

Trong nhiều trường hợp có những Case không quan sát được, hay không thu thập được số liệu cho một biến nào đó. Nhiều trường hợp do khi chỉ định vào số liệu ta không chỉ định giá trị thiếu của biến trong trường hợp này SPSS dùng dấu chấm (.) kí hiệu giá trị thiếu. Có thể sử dụng thủ tục Recode nói ở trên với việc chọn nhóm theo cùng biến cũ để đặt bổ sung các giá trị này, các giá trị được bổ sung như vậy thường là các giá trị cá biệt không có tác dụng xử lý số liệu sau này. Trong trường hợp ta muốn các giá trị thiếu có nội dung để sử dụng trong các phân tích sau này cần chỉ ra cách xác định các giá trị đó. Có thể dùng các thủ tục khác nhau để lấp đầy dữ số liệu, SPSS cho phép chọn 1 trong 5 cách khác nhau để tạo giá trị thay cho các giá trị thiếu.

Với chỉ định Transform/ Replace missing values ta có bảng:

Với các phương pháp khác nhau ta tạo ra các biến mới trong đó bảo tồn các giá trị quan sát được và thêm vào vị trí thiếu các giá trị mà từng phương pháp tạo ra. Hãy thực hiện việc xác định các giá trị thiếu cho biến q09 (thu thập sau 5 năm vay vốn) bằng cách chọn các phương pháp khác nhau. Đơn giản nhất là lấy giá trị trung bình của tất cả các quan sát, hoặc giá trị trung vị của phân phối với các quan sát xung quanh giá trị thiếu.



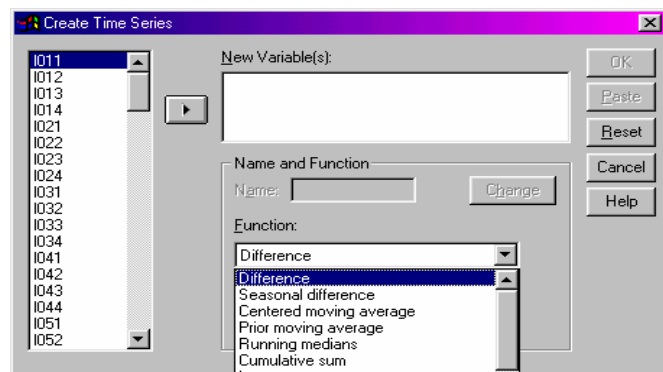
* Chọn số liệu để phân tích: Data/ Select case

Trong nhiều trường hợp ta cần phân tích, mô tả một số biến chỉ đối với các cá thể (case) thỏa mãn một số điều kiện nào đó. Sử dụng chỉ định **Data/ Select case** ta có thể chọn tập số liệu theo điều kiện của một biến nào đó. Nếu muốn thu hẹp hơn nữa phạm vi có thể sử dụng tiếp chỉ định này.

* Tạo chuỗi thời gian: Transform/ Create Time Series

Đối với các bộ số liệu quan sát theo trình tự thời gian, khi phân tích ta thường quan tâm nhiều đến tính ổn định, quán tính, ... của quá trình tương ứng.

Thủ tục này cho phép tạo một chuỗi thời gian theo một vài cách khác nhau. Khi chọn một phương pháp nào đó, cần chú ý đến ý nghĩa của biến mới tạo ra. Có thể đặt biến theo sở thích, ngầm định biến sẽ được đặt tên bằng cách nối vào cuối tên biến cũ hai ký tự (_1). Lệnh chỉ định: Transform/Create Time series:



* Tạo tệp giá trị trung bình nhóm: Data/ Aggregate

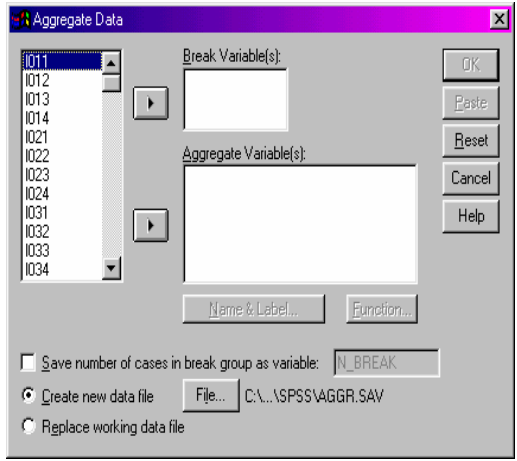
Thủ tục Aggregate cho phép tạo một tệp trong đó phân nhóm theo giá trị của một số biến nào đó. Đây là một trong những thủ tục rất tiện dụng trong các bước tổng hợp số liệu thống kê. Các biến mới được tạo ra nhờ các hàm của từng biến theo ý muốn của người sử dụng. Chú ý rằng nếu trong nhóm có giá trị thiếu hệ thống thì giá trị trung bình sẽ không được tính. Muốn khắc phục hiện tượng này cần chọn các Case ở đó giá trị của biến trong công thức không bị bỏ trống (nhờ thủ tục Select case ... đã nói ở trên). Cũng cần chú ý đến hai tự chọn cuối cùng là ghi với tên tệp mới hay ghi đè lên tệp cũ. Trong đa số các trường hợp nên ghi thành 1 tệp với tên tệp mới. Khi chỉ định Data/ Aggregate:

Chọn một hay một số biến để phân nhóm ghi vào ô Break variable

Chọn biến hoặc hàm của ác biến khác ghi vào ô Aggregate variable (có thể tự đặt hàm hay chọn hàm nhờ nút Function. Hàm mặc định là giá trị trung bình, vì vậy, ta có thể chỉ cần chọn biến lúc đó biến mới sẽ có phần đầu của tên biến là tên biến cũ với hai kí tự bổ sung là _1. Thí dụ, với tệp Vay.sav ta có muốn tạo một tệp trong đó có tuổi chủ hộ (v02) và thu nhập trung bình của các hộ có chủ hộ cùng tuổi ta chọn v02 cho Break variables và q09 (thu nhập sau vay vốn) cho Aggregate, đặt tên tệp mới là aggr1.sav, đồng thời chọn biến n_break thì một tệp mới có tên Aggr1.sav chứa ba biến v02, q09_1 và n_break. Giá trị của các biến được xác định như sau: $q09_1 = \text{mean}(q09/v02)$, chúng ta có thể đặt ngay nhãn và giá trị nhãn nếu cần cho các biến mới, chẳng hạn đặt nhãn cho q09_1 là thu nhập TB hộ sau vay vốn theo tuổi chủ hộ; n_break là số quan sát có cùng giá trị của biến v02.

Thủ tục tạo tệp này cho phép tạo tệp với các biến mới không chỉ là các giá trị trung bình. Có thể sử dụng Function để chọn các giá trị đặc trưng thống kê khác như độ lệch tiêu chuẩn mẫu, giá trị lớn nhất, nhỏ nhất, các tỷ lệ theo từng giá trị của biến, ... Với thủ tục này, các chỉ tiêu thống kê đơn giản được tính toán dễ dàng, tuy vậy có nhiều lựa chọn trên thực đơn không liệt kê hết, vì vậy chúng ta sẽ xem xét kỹ hơn trong phần ngôn ngữ lập trình từ SPSS.

Nếu mở tệp Aggr1.sav ta có:



The 'Aggregate Data' dialog box shows 'v02' selected for 'Break Variable(s)' and 'q09' for 'Aggregate Variable(s)'. The 'Save number of cases in break group as variable' checkbox is checked, with 'N_BREAK' entered. The 'Create new data file' option is selected, with the file path 'C:\...\SPSS\AGGR.SAV'.

	v02	q09_1	n_break	vaf	vaf	vaf	vaf
1	26	3162.50	1				
2	28	11025.00	2				
3	29	23610.00	2				
4	30	6183.00	7				
5	31	6007.50	2				
6	32	8629.00	6				
7	35	8656.00	5				
8	36	7475.00	5				
9	37	14269.20	6				
10	38	6083.33	3				
11	39	.	1				

* Bảng

Bảng có lẽ là một trong những công cụ thông dụng nhất để mô tả và tổng kết dữ liệu của bạn. Bên cạnh đó, rất nhiều kết xuất của SPSS được thể hiện dưới dạng bảng trụ (pivot table) đã sử dụng thêm tầm quan trọng của việc cần phải hiểu biết kỹ thuật lập và hiệu đính các bảng.

* Một số khái niệm cơ bản thường gặp

- Các biến:

+ Biến phân nhóm (categorical variables)

Loại này thường có một vài trị số, ví dụ như giới tính, vùng, .. các biến được phân nhóm theo mục đích phân loại và bạn thường gặp chúng qua các giá trị biến hoặc nhãn giá trị biến được xuất hiện theo cột (chủ từ) hoặc theo hàng (tân từ) trong một bảng.

+ **Biến tổng kết (summary variables)**

Loại này cung cấp nội dung (những thống kê) của các ô trong bảng.

+ Ngoài ra, trong SPSS còn có một loại biến thứ ba gọi là các tập hợp đa lựa chọn (multi response sets) mà chúng ta sẽ không đề cập trong cuốn sách này.

- **Cấu trúc của bảng:**

Trong SPSS các bảng được thiết kế thông qua các chiều.

Có 3 loại chiều:

+ **Chiều theo hàng (row dimension)**

+ **Chiều theo cột (column dimension)**

+ **Chiều theo lớp/ trang (layer dimension)**

Chúng được định nghĩa bởi các biến hàng, biến cột và biến lớp. Phần chính của bảng được tạo nên bởi các ô có chứa các thống kê.

Trong cùng chiều thì các biến lại có thể được xếp chồng (stacked) hoặc xếp lồng (nested). Các biến được xếp chồng thể hiện riêng rẽ trong khi nếu chúng được xếp lồng thì các nhóm của biến xếp lồng sẽ được thể hiện cho từng nhóm của biến ở trên nó. (Tức là trong từng cột, từng hàng lại được phân tổ nhỏ - Bảng thống kê phức tạp).

- **Bảng thống kê theo mặc định (Default table)**

Trong số các loại bảng SPSS bạn có thể chọn các bảng cơ bản (basic tables), các bảng tổng quát (general tables) và các bảng tần số (tables of frequencies). Các bảng cơ bản có thể đáp ứng đầy đủ cho hầu hết cá mục đích. Các bảng tổng quát thường được dùng cho các bảng phức tạp khi việc xếp chồng, xếp lồng, hoặc các thống kê khác nhau được áp dụng cho các biến khác nhau.

* **Xác định các bảng giả định (dummy tables)**

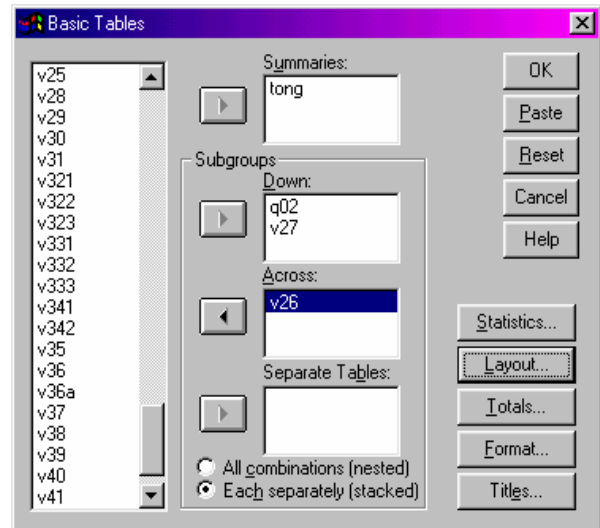
Trước khi tiến hành lập một bảng, bạn nên bắt đầu việc định nghĩa bảng giả định cho nó. Bảng giả định là một bảng với các ô không có nội dung (dữ liệu) gì nhưng chủ yếu để nói rằng những biến nào sẽ được sử dụng theo chiều nào của bảng, loại số liệu thống kê nào cần được tính toán, hoặc định dạng của bảng ra sao v.v.. Có một bảng giả định trong tay sẽ thuận tiện rất nhiều khi bạn lựa chọn các tùy chọn trong những hộp thoại của thủ tục lập bảng của SPSS.

Lập 1 bảng giả định có thể rất đơn giản, đôi khi chỉ cần 1 chiếc bút chì phác thảo trên một tờ giấy. Bạn cũng có thể sử dụng định dạng từ các nguồn khác, ví dụ như báo cáo về các hoạt động của hộ gia đình hoặc từ 1 cuốn niên giám thống kê.

* **Các bảng cơ bản (Basic tables)**

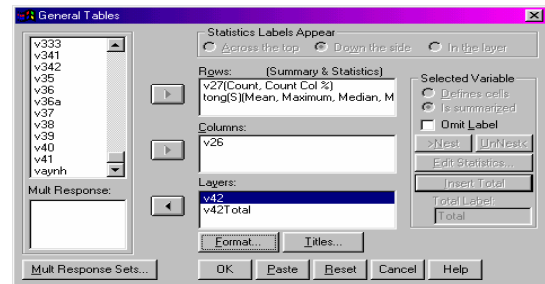
Thủ tục Basic Tables tables lập các bảng chéo thể hiện các thống kê cho các nhóm phụ. Những biến tổng kết sẽ được tổng hợp trong các ô được xác định bởi các biến lập nhóm phụ (subgroup variable). Các thống kê giống nhau sẽ được báo cáo cho mọi biến được tổng kết ở trong bảng. Thống kê mặc định được trình bày trong bảng là số trung bình. Nếu không có biến tổng kết nào, các tần số sẽ được tổng hợp trong các ô.

Để thực hiện nhiệm vụ này từ thanh menu, hãy chọn Statistics/ Custom Tables/ Basic Tables ..



* Các bảng tổng hợp (General tables)

Nếu bạn có ý định tính toán các thống kê khác nhau cho các biến khác nhau trong cùng một chiều của bảng, hoặc nếu bạn muốn kết hợp xếp chồng và xếp lồng cũng như tính toán các tổng số phức tạp trong một bảng, bạn phải chọn thủ tục: Statistics/ Custom Tables/ General Tables



* Hiệu chỉnh các bảng trụ

Như đã nêu ở trên, hầu hết cá thủ tục của SPSS đều tạo ra các bảng trụ/ bảng xoay (pivot table) là những kết xuất mà bạn có thể hiệu chỉnh chúng sao cho thích hợp nhất với mục đích phân tích của bạn.

Để hiệu chỉnh một bảng, trước tiên nhấp đúp bất kỳ vị trí nào trong bảng để kích hoạt nó. Để ẩn dấu một cột hay dòng, nhấn đồng thời (Ctrl + Alt + nhấp chuột) lên nhãn của cột hay dòng và sau đó chọn Hide từ menu View. (Một cách làm tắt là nhấp chuột phải lên nhãn và chọn Select/ Data Cell and Label/ nhấp phím phải chuột/ Hide Category). Để thay đổi độ rộng của cột, kéo cạnh bên phải của nó trong vòng khu vực của nhãn. Để thay các chiều của bảng, chọn Pivoting Trays (các khay trụ) từ menu Pivot và kéo rê một trong những biểu tượng của nó từ một khay đến khay khác, hoặc đảo vị trí các biểu tượng trong một chiều (đối với SPSS 8.0 thì có thêm thủ tục Statistics/ Summaries/ Layered Reports; đối với SPSS 9.0 thì thủ tục Analyze/ Reports/ OLAP Cube.. sẽ thay thế cho việc chọn Pivoting Trays).

Cách tốt nhất là để khám phá trình hiệu chỉnh bằng trụ Pivot Table Editor là hãy mở nó ra và tự tạo ra kinh nghiệm cho bản thân mình.

* Các báo cáo tổng kết (Statistic/ Summarize/ Reports ..)

Trong SPSS bạn có thể dễ dàng tạo ra các báo cáo tổng kết. Dưới góc độ báo cáo tổng kết có 2 loại biến: các biến báo cáo (report variable) và các biến ngắt (break variable). Các biến báo cáo thường xuất hiện trong các cột dữ liệu. Các biến ngắt chia dữ liệu thành các nhóm được thể hiện trong các cột chủ từ (break column).

- Báo cáo tổng kết theo cột (Report Summaries in Columns)

Trình bày các thống kê mô tả mà không liệt kê danh sách các chủ đề.

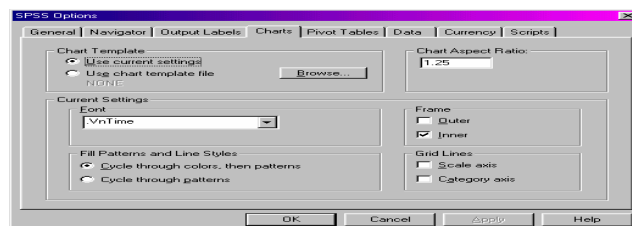
- Báo cáo tổng quát theo hàng (Report Summaries in Rows)

Cũng được tạo theo cách tương tự như Report Summaries in Columns. Sự khác biệt là ở chỗ các thống kê sẽ xuất hiện trong các hàng và do vậy bạn có thể lập danh sách các trường hợp riêng rẽ.

* Biểu đồ/ đồ thị (Charts): Edit/ Option

Thông thường bạn muốn nâng cao chất lượng báo cáo của mình bằng cách đưa chúng vào các biểu đồ/ đồ thị khác nhau. SPSS cung cấp một số lượng lớn các dạng mẫu đồ thị: trong số đó thì đồ thị thanh (bar chart) và đồ thị tròn (pie charts) có lẽ là được sử dụng thường xuyên nhất. Tất cả các đồ thị có thể được hiệu chỉnh sau khi chúng đã được lập. Các tiện ích hiệu chỉnh được áp dụng giống nhau cho mọi dạng đồ thị khác nhau. Ngoài ra trong SPSS còn cung cấp một dạng đồ thị tương tác (interactive graph).

Trước tiên để có thể sử dụng phông tiếng Việt được trong khi hiệu chỉnh đồ thị, từ thanh menu ở cửa sổ SPSS Data Editor, hãy chọn Edit/ Option



Sau khi chọn được phông tiếng Việt, tiếp tục từ menu ở cửa sổ SPSS Data Editor, hãy chọn Graphs/ Chọn dạng đồ thị (Bar.. , Line)

Để hiệu chỉnh hãy bắt đầu bằng nhấp đúp vào đồ thị và cửa sổ hiệu chỉnh đồ thị SPSS Chart Editor sẽ cho phép chúng ta lựa chọn.

* In dữ liệu: File/ Print

Đôi khi bạn muốn in một số tập hợp nhỏ các trường hợp (ví dụ như 30 trường hợp đầu tiên) cùng với một vài biến đáng quan tâm nào đó. Cách dễ dàng nhất là chọn một vùng cần thiết trong cửa sổ SPSS Data Editor, sau đó sử dụng thủ tục File/ Print .. Kết xuất sẽ được gửi thẳng đến máy in theo một cách thức được định dạng chuẩn. Nhược điểm của chọn vùng là các biến phải nằm kế liên tục nhau trong SPSS Data Editor.

Trong thực tế các biến cần in ra lại thường không nằm kế tiếp nhau trong cửa sổ SPSS Data Editor. Để chọn được các biến như ý muốn, hãy sử dụng thủ tục Statistics/ Summarize/ Case Summaries..

Chương II. ỨNG DỤNG PHẦN MỀM EXCEL NÂNG CAO TRONG KINH TẾ VÀ QUẢN TRỊ KINH DOANH

1. Các hàm ứng dụng trong EXCEL

1.1. Các hàm khấu hao

Thí dụ 1: Có tài liệu về một số tài sản cố định của một Công ty như sau.

Tên tài sản	Nguyên giá (tr.đ)	Giá trị thu hồi (triệu đ)	Số năm sử dụng (năm)
A	150	10	8
B	145	8	8
C	120	8	5
D	750	20	10
E	150	10	5
F	350	15	10
H	80	5	5

Yêu cầu: Tính khấu hao hàng năm cho từng tài sản cố định và số tiền khấu hao hàng năm của cả công ty?

Trong tài chính có nhiều phương pháp khấu hao với các công thức tính đã được xây dựng (khấu hao theo thời gian sử dụng, khấu hao theo khối lượng công việc, khấu hao giảm dần và khấu hao trả dần vốn đầu tư). Dựa vào các công thức tính này, các nhà tin học đã thiết lập các hàm khấu hao trong EXCEL như sau:

a) Hàm SLN (Straight - line depreciation)

Đây là hàm khấu hao theo phương pháp đường thẳng (khấu hao theo thời gian) của 1 tài sản cố định cho 1 thời kỳ.

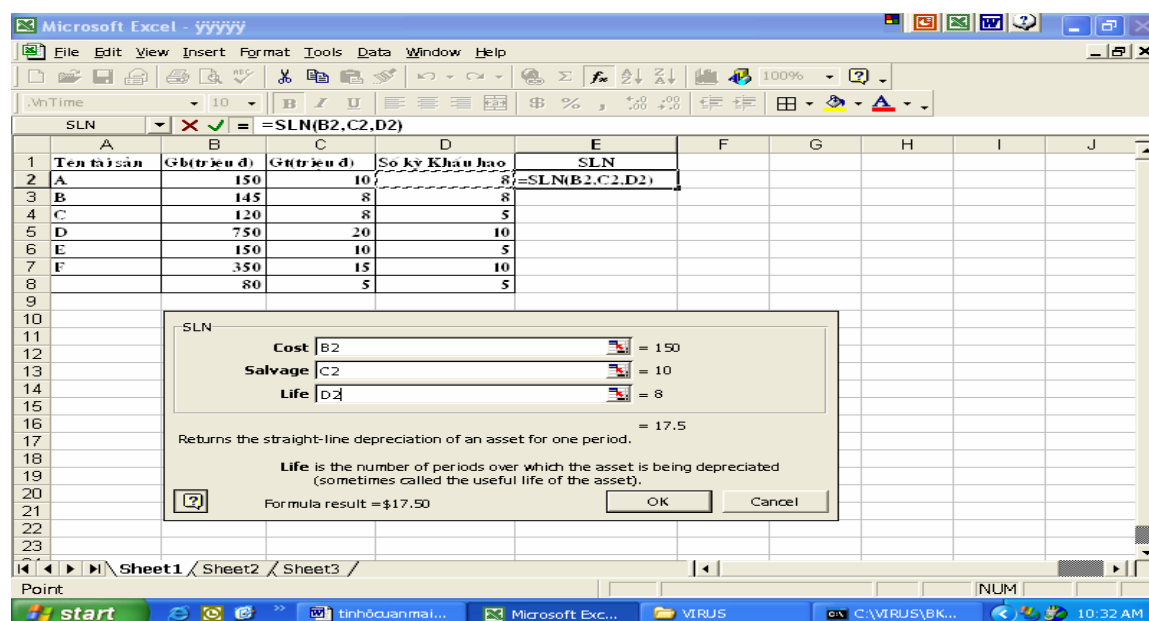
Cú pháp: =SLN(Cost,Salvage,Life)↴

Trong đó: Cost là giá trị nguyên giá của tài sản cố định.

Salvage là giá trị thu hồi khi thanh lý của tài sản.

Life là thời kỳ khấu hao (tuổi thọ sử dụng tài sản).

Sử dụng dữ liệu ở thí dụ 1, chúng ta có thể tính khấu hao theo các thao tác sau:



- Thiết kế cơ sở dữ liệu trên EXCEL (xem hình trên)
- Chọn Insert/Function/Financial/SLN
- Khai báo địa chỉ ô cho Cost, Salvage, Life

- OK

b) Hàm DB (Declining Balance)

Đây là hàm khấu hao giảm dần tính theo phương pháp số dư giảm dần trong một khoảng thời gian xác định.

Cú pháp: =DB(Cost,Salvage, Life,Period,Month)↵

Trong đó: Cost là giá trị nguyên giá của tài sản.

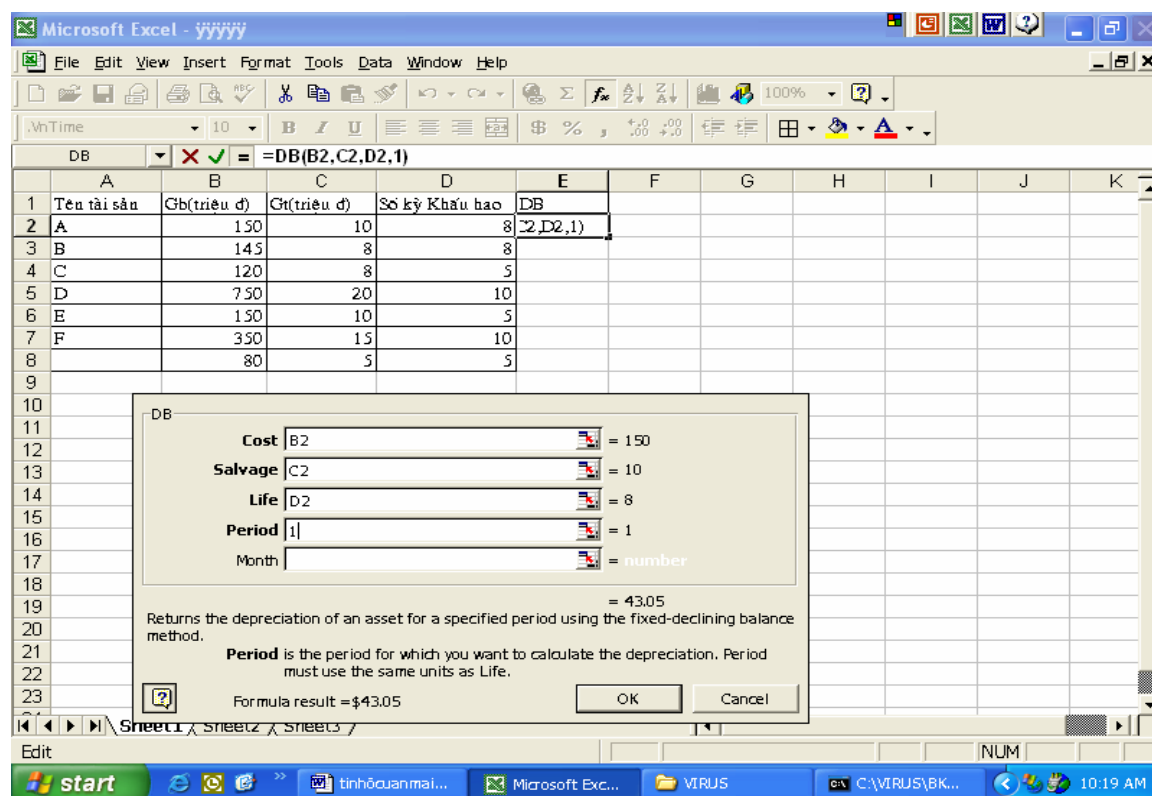
Salvage là giá trị thu hồi khi thanh lý của tài sản.

Life là tuổi thọ sử dụng của tài sản (số năm, số tháng...).

Period là kỳ khấu hao (thứ 1,2,3,4,... n; kỳ phải cùng đơn vị đo với Life).

Month là số tháng sử dụng trong năm đầu, nếu không ghi máy mặc nhận = 12.

Thí dụ: Sử dụng dữ liệu ở thí dụ 1, chúng ta tính khấu hao năm thứ 1 theo hàm DB như sau:



- Thiết kế cơ sở dữ liệu trên EXCEL (xem hình trên)

- Chọn Insert/Function/Financial/DB

- Khai báo địa chỉ ô cho Cost, Salvage, Life, Period, Month (nếu có) cho năm thứ nhất

- OK

Tương tự như vậy ta tính DB cho năm thứ 2,3,4...

c) Hàm DDB (Double-declining balance)

Đây là hàm khấu hao theo phương pháp số dư giảm dần gấp đôi (giảm nhanh kép) cho một tài sản trong một khoảng thời gian xác định.

Cú pháp: =DDB(Cost,Salvage,Life,Period,Factor)↵

Trong đó: Cost là giá trị nguyên giá của tài sản.

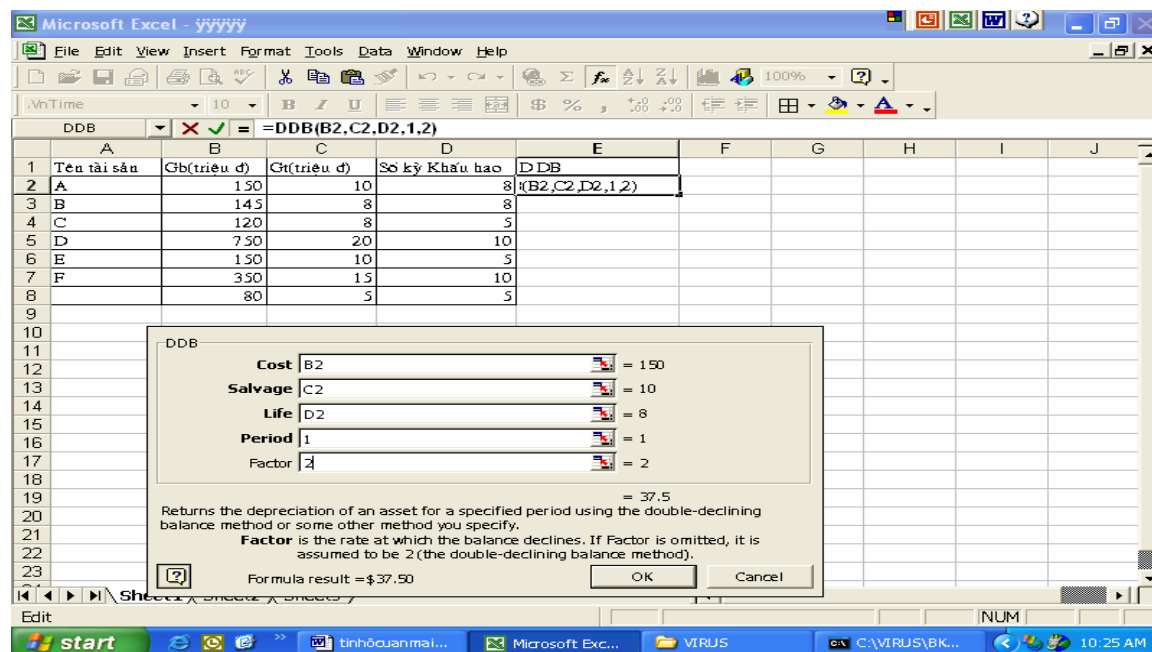
Salvage là giá trị thanh lý tài sản.

Life là tuổi thọ sử dụng của tài sản (số năm, số tháng...).

Period là kỳ khấu hao (thứ 1,2,3,4,...n; kỳ phải cùng đơn vị đo với Life).

Factor là hệ số giảm nhanh, thường căn cứ tuổi thọ sử dụng của tài sản mà qui định được tính là 1,5 hoặc 2 (tức là gấp 1,5 hoặc 2 lần) mức khấu hao đều. Nếu không ghi Factor thì máy mặc định bằng 2 (tức 200%) (phương pháp giảm nhanh kép).

Thí dụ: Vẫn lấy thí dụ 1 trên, hãy tính giá trị khấu hao cho năm đầu tiên với hệ số khấu hao = 2. Thao tác tính toán trên máy như sau:



- Thiết kế cơ sở dữ liệu trên EXCEL (xem hình trên)
- Chọn Insert/Function/Financial/DDB
- Khai báo địa chỉ ô cho Cost, Salvage, Life, Factor
- OK

Tương tự ta tính khấu hao cho các năm sau với hệ số khấu hao bằng 2.

d) Hàm SYD (Sum of year digits)

Đây là hàm khấu hao giảm dần được tính theo phương pháp tổng các số thứ tự năm sử dụng tài sản (phương pháp tổng số số năm sử dụng) cho một tài sản trong một khoảng thời gian xác định.

Cú pháp: =SYD(Cost,Salvage,Life,Per)↵

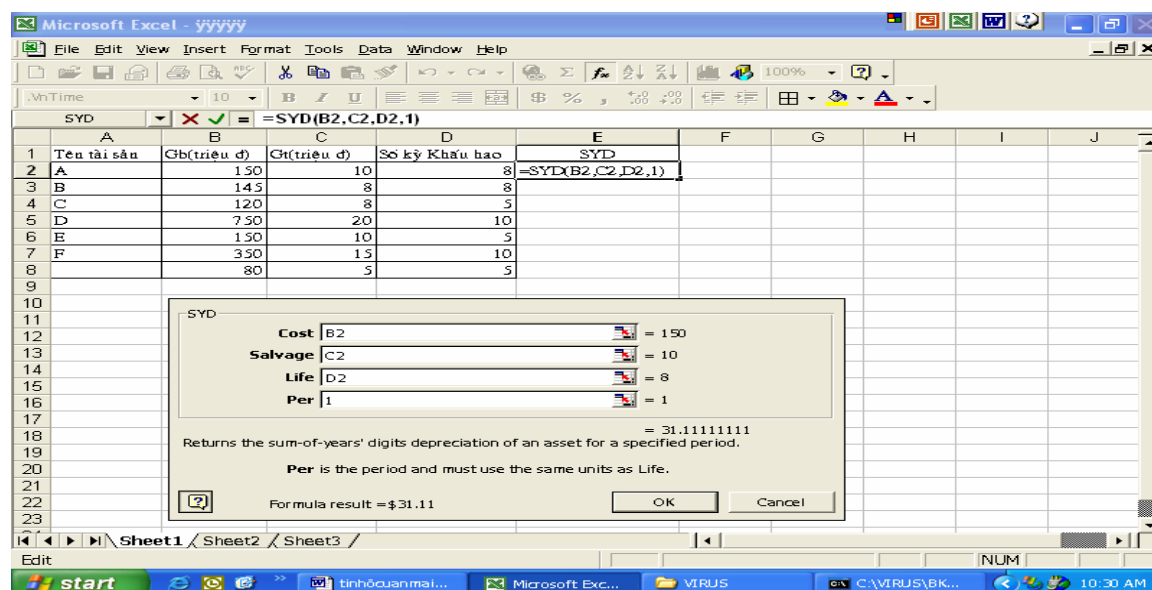
Trong đó: Cost là giá trị nguyên giá của tài sản.

Salvage là giá trị thu hồi khi thanh lý của tài sản.

Life là tuổi thọ sử dụng tài sản (số năm, tháng...).

Per là số thứ tự năm khấu hao (phải cùng đơn vị với Life)

Thí dụ: Lấy dữ liệu ở thí dụ 1 trên đây tính khấu hao năm đầu tiên theo hàm SYD trên máy tính như sau:



- Thiết kế cơ sở dữ liệu trên EXCEL (xem hình trên)
- Chọn Insert/Function/Financial/SYD
- Khai báo địa chỉ ô cho Cost, Salvage, Life, Per
- OK

Tương tự ta tính khấu hao cho các năm sau.

1.2 Các hàm tính toán hiệu quả vốn đầu tư

Khi đánh giá dự án đầu tư, chúng ta thường nhận thấy chúng có một đặc điểm chung là lợi ích và chi phí của các dự án thường xuất hiện ở các thời điểm khác nhau. Xuất phát từ quan điểm “1 đồng hôm nay có giá trị cao hơn 1 đồng nhận được vào ngày mai”, bởi vì đồng tiền có được ngày hôm nay có thể được đem đầu tư và sinh lời ngay lập tức, các chi phí và lợi ích đến sớm hơn về mặt thời gian cần phải được đánh giá cao hơn và ngược lại. Trong thực tế, người đi vay (cần vốn) sẵn sàng trả lãi để có quyền sử dụng số vốn vay đó, còn người cho vay (dư vốn) đòi hỏi phải được trả lãi.

Như vậy, tiền tệ có giá trị theo thời gian. Nếu đem đầu tư 1 đồng ngày hôm nay với lãi suất (r), thì giá trị tiền sau 1 năm, 2 năm và t năm được tính như sau:

Diễn giải	Sau 1 năm	Sau 2 năm		Sau t năm
Giá trị tương lai	$P_1 = 1 * (1+r)$	$P_2 = 1 * (1+r)^2$		$P_t = 1 * (1+r)^t$
Giá trị tương lai qui về hiện tại	$Po = P_1 / (1+r)$	$Po = P_1 / (1+r)^2$		$Po = P_t / (1+r)^t$

Để đánh giá hiệu quả của dự án đầu tư thông qua việc so sánh lợi ích đạt được so với chi phí bỏ ra, thì giá trị lợi ích và chi phí bỏ ra ở các thời điểm khác nhau phải được đưa về cùng một thời điểm để so sánh. Trong tài chính chúng ta đã có 2 phương pháp để qui đổi giá trị của tiền đầu tư gắn liền với 2 thời điểm là hiện tại và tương lai thông qua suất chiết khấu (r). Phù hợp với cách tính này các nhà tin học đã lập trình các hàm đánh giá hiệu quả đầu tư trên EXCEL như sau:

a) Hàm FV (Future value) - Giá trị tương lai

Đây là hàm tính giá trị tương lai của 1 đầu tư trong một khoảng thời gian ứng với mức lãi suất xác định.

Thí dụ 2: Có một số khách hàng gửi tiền vào ngân hàng nông nghiệp & PTNT với lãi suất hàng năm, số kỳ rút tiền và số tiền gửi như sau. Hãy tính số tiền mà họ còn ở ngân hàng sau 1 năm?

Khách hàng	Số tiền gửi	Lãi suất năm	Số kỳ rút lãi	Số tiền rút mỗi
------------	-------------	--------------	---------------	-----------------

	(USD)	(%)	(lần)	kỳ (USD)
A	20000	1.5	4	1000
B	25000	2.0	4	1000
C	10000	1.5	4	500
D	15000	2.0	4	1000
E	15000	2.0	4	500
F	15000	1.6	4	600
H	10000	1.8	4	500

Sử dụng hàm này tính giá trị tương lai của một khoản tiền đầu tư thường gặp ở các trường hợp sau:

* **Tính FV khi thanh toán định kỳ, lãi suất ổn định**

Cú pháp: =FV(Rate,Nper,Pmt,Pv,Type)

Trong đó:

Rate là lãi suất/ kỳ

Nper là số kỳ trả tiền

Pmt là tiền thanh toán mỗi kỳ.

Pv là giá trị tiền hiện tại, nếu không ghi thì máy mặc nhận = 0.

Type là giá trị số bằng 0 hay 1, chỉ định phương thức trả tiền trong kỳ.

Type	Thời gian
0	Trả cuối kỳ
1	Trả lại đầu chu kỳ

The screenshot shows a Microsoft Excel spreadsheet with the following data:

	A	B	C	D	E	F
1			Số kỳ rút tiền	Số tiền rút mỗi kỳ	Số tiền gửi	FV
2	Chủ đầu tư	Rate		(USD)	(USD)	
3	A	0.015	4	1000	20000	=FV(B3,C3,D3,-E3,0)
4	B	0.020	4	1000	25000	
5	C	0.015	4	500	10000	
6	D	0.020	4	1000	15000	
7	E	0.020	4	500	15000	
8	F	0.016	4	600	15000	
9	H	0.018	4	500	10000	

The FV dialog box is open, showing the following values:

- Rate: B3 = 0.015
- Nper: C3 = 4
- Pmt: D3 = 1000
- Pv: -E3 = -20000
- Type: 0 = 0

The formula result is 17136.36764.

Chú ý: Khi sử dụng các hàm tính toán giá trị của khoản tiền đầu tư cần xác định và biểu thị dòng tiền. Thông thường có 2 trường hợp cần xác định và biểu thị dòng tiền như sau:

- Dòng tiền mang dấu âm (-) được xác định khi: Đầu tư mua trang thiết bị SXKD, cho vay, gửi vào ngân hàng, mua chứng khoán hoặc đi vay về tiếp tục đầu tư vào SXKD, tiền thanh toán định kỳ khi đi vay (được tính vào thời điểm hết hạn vay), tiền được thanh toán định kỳ khi cho vay (ước tính tại thời điểm bắt đầu cho vay), (tiền ra khỏi túi ông chủ).
- Dòng tiền mang dấu dương (+) được xác định khi: Đi vay về chưa đầu tư vào SXKD, tiền rút từ ngân hàng về, tiền nhận được từ các khoản thanh toán với khách hàng, tiền được thanh toán định kỳ khi cho vay (được tính vào thời điểm hết hạn vay) ... (tiền vào túi ông chủ, tiền ước tính vào túi ông chủ).

* **Tính FV khi thanh toán 1 lần cả vốn và lãi (lãi gộp) với lãi suất cố định**

$$FV = PV \cdot (1 + \text{Rate})^{\text{Life}}$$

Trong đó: Rate là lãi suất/ kỳ

Pv là giá trị tiền hiện tại

Life là thời gian đầu tư (tháng, năm), phù hợp với kỳ tính Rate.

Thí dụ: Lấy lại thí dụ 2 trên. Hãy tính số tiền mà họ được nhận từ ngân hàng sau 1 năm, 2 năm và 3 năm?

	A	B	C	D	E	F
	Chủ đầu tư	Rate	Số kỳ rút tiền	Số tiền rút mỗi kỳ (USD)	Số tiền gửi (USD)	FV
3	A	0.015	4	1000	20000	=e3(1+b3)^1
4	B	0.020	4	1000	25000	
5	C	0.015	4	500	10000	
6	D	0.020	4	1000	15000	
7	E	0.020	4	500	15000	
8	F	0.016	4	600	15000	
9	H	0.018	4	500	10000	

* **Tính FV khi thanh toán 1 lần cả vốn và lãi, với lãi suất thay đổi**

Thí dụ 3: Có tài liệu sau đây

Khách hàng	Số tiền gửi (USD)	Lãi suất năm(%)		
		Năm thứ 1	Năm thứ 2	Năm thứ 3
A	20000	1.5	2.0	1.5
B	25000	2.0	2.5	2.0
C	10000	1.5	2.0	1.5
D	15000	2.0	2.2	2.2
E	15000	2.0	2.0	2.1
F	15000	1.6	2.1	1.8
H	10000	1.8	2.5	1.6

Hãy tính số tiền mà họ được nhận từ ngân hàng sau 3 năm?

Trong trường hợp này khi tính FV ta sử dụng hàm FVSCCHEDULE

Cú pháp: = FVSCCHEDULE(Principal, Schedule)

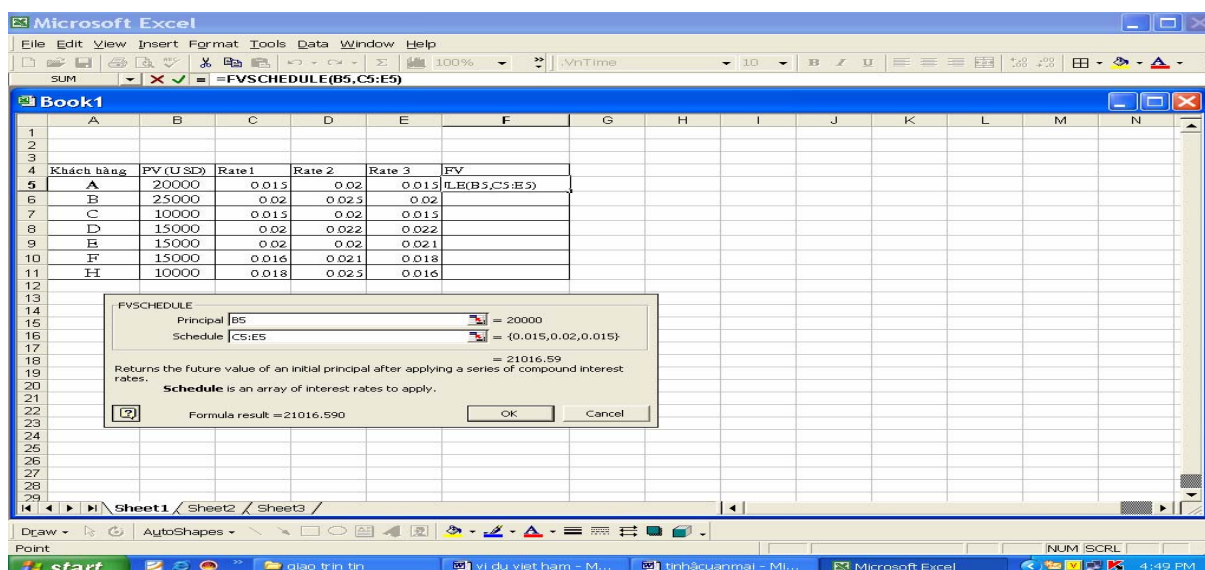
Hoặc: **FVSCCHEDULE = Principal * (1+Rate₁) * (1+Rate₂) * ... (1+Rate_i)**

Trong đó: Principal là giá trị ban đầu của 1 đầu tư

Schedule là 1 dãy tỷ lệ lãi suất được áp dụng

Rate_i là lãi suất kỳ thứ i, (với i = 1, n)

Sử dụng số liệu ở thí dụ 3, ta giải trên EXCEL như sau:



b) Hàm PV (Present Value) - Giá trị thực tại hoá

Đây là hàm tính giá trị khoản tiền ở một thời điểm nào đó quy về thời điểm thực tại theo lãi suất chiết khấu nhất định (thường gọi là giá trị thực tại hoá) của một khoản tiền đầu tư.

Thí dụ: Một nhà sản xuất có cơ hội được đầu tư một số tiền là 50 triệu với lời hứa sẽ nhận được 57 triệu sau đó 1 năm. Anh ta tính toán và trả lời “Tôi từ chối đầu tư này bởi vì tôi chỉ có 10 triệu là vốn tự có của tôi, còn tôi phải đi vay 40 triệu với lãi suất 10%/năm. Vì vậy, tôi chỉ thu được 53 triệu. Tôi không thể hài lòng vì đầu tư này chỉ đạt được tỷ suất là 6%/năm, trong khi tôi phải đi vay với lãi suất 10%/năm”

Anh chỉ hãy cho đánh giá của mình về câu trả lời trên của nhà đầu tư về:

- Quyết định trên của anh ta có đúng không?

- Hãy giải thích bằng phân tích cụ thể?

Trong tài chính sử dụng các công thức tính PV ta có thể đánh giá rằng quyết định của nhà sản xuất trên là hoàn toàn đúng. Nhưng bằng hàm FV đã được lập trình trên EXCEL chúng ta tính toán và có nhận xét đúng như trên.

Khi tính PV ta cũng thường gặp ở các trường hợp sau:

* Tính PV khi thanh toán định kỳ, lãi suất ổn định

Cú pháp: =PV(Rate,Nper,Pmt,Fv,Type)↵

Trong đó: Rate là lãi suất/kỳ.

Nper là số kỳ trả tiền.

Pmt là giá trị thanh toán mỗi kỳ.

Fv là giá trị khoản tiền tương lai, nếu không ghi Fv thì nó được coi = 0.

Type là chỉ định phương thức trả tiền nhân 2 giá trị số = 0 hay 1.

Type	Thời gian
0	Trả cuối kỳ
1	Trả lại đầu chu kỳ

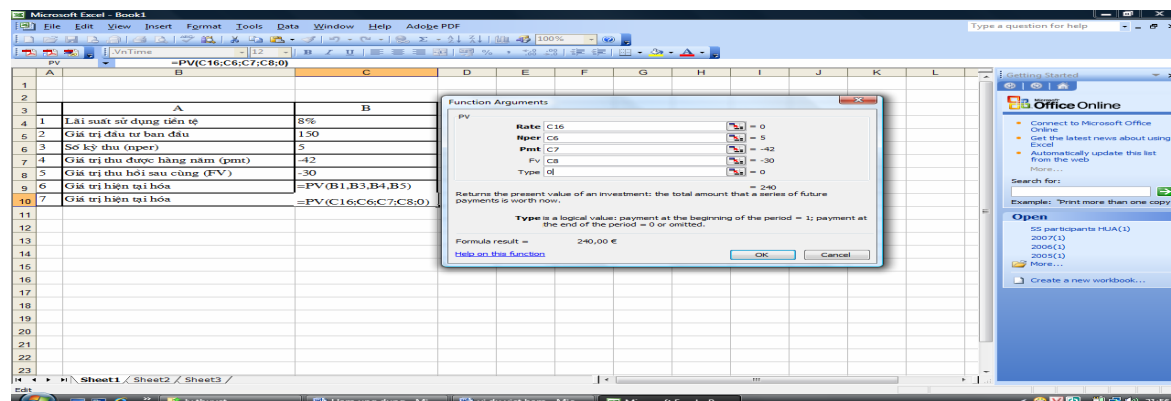
Thí dụ 4: Một nhà máy dự kiến trang bị thêm một thiết bị máy ép nhựa với giá mua và lắp đặt là 150 triệu. Theo tính toán hàng năm, giá trị ròng do sử dụng thiết bị mới sản xuất mang lại là 42 triệu (đã trừ chi phí sản xuất). Dự án dài 5 năm. Sau khi thanh lý, giá trị thu hồi máy móc là 30 triệu. Với lãi suất sử dụng tiền tệ là 8%/năm, anh (chị) khuyên nhà máy nên đầu tư hay không nên? vì sao?

Lập bảng số liệu trên EXCEL

	A	B
1	Lãi suất sử dụng tiền tệ	8%

2	Giá trị đầu tư ban đầu	150
3	Số kỳ thu (nper)	5
4	Giá trị thu được hàng năm (pmt)	-42
5	Giá trị thu hồi sau cùng (FV)	-30
6	Giá trị hiện tại hóa	=PV(B1,B3,B4,B5)
7		

Giải trên E XCEL có kết quả sau.



Chú ý: (i) Nếu giá trị thu hàng năm không đều, tính PV từng năm sau đó cộng lại thành PV của cả dự án, khi đó Pmt hàng năm coi như bằng 0.

(ii) Trong EXCEL các nhà tin học không lập trình tính hàm PV khi thanh toán 1 lần cả vốn và lãi với lãi suất không thay đổi và lãi suất thay đổi, nhưng từ công thức tính hàm FV ở hai trường hợp này chúng ta suy ra các công thức tính PV như sau:

*** Tính PV khi thanh toán 1 lần cả vốn và lãi (lãi gộp)**

$$PV = FV / (1 + Rate)^{Life}$$

Trong đó: Life là thời gian đầu tư (tháng, năm), phù hợp với kỳ tính Rate.

*** Tính PV khi thanh toán 1 lần cả vốn và lãi, với lãi suất thay đổi**

$$Principal = FV / SCHEDULE / (1 + Rate_1) * (1 + Rate_2) * \dots * (1 + Rate_n)$$

Với công thức này hoàn toàn tính toán trên bảng tính E XCEL.

c) Hàm NPV (Net Present Value)- Giá trị thực tại hóa ròng

Đây là hàm tính giá trị thực tại hóa ròng của một giá trị đầu tư dựa trên 1 loạt các dòng tiền thu, chi tài chính theo định kỳ với 1 lãi suất chiết khấu nhất định. Giá trị thực tại hóa ròng của 1 đầu tư là giá trị hôm nay của 1 loạt các thanh toán (âm) và thu nhập (dương) trong tương lai.

Đối với dự án đầu tư: Là sự chênh lệch giữa giá trị hiện tại của khoản tiền thu được (lợi ích của dự án) và giá trị hiện tại của dòng tiền chi (chi phí của dự án) của 1 dự án đầu tư.

Vì dự án đầu tư thường thực hiện trong nhiều năm, nên các khoản thu và chi của dự án được rải ra trong nhiều năm, do đó muốn so sánh lợi ích và chi phí của dự án một cách chính xác chúng ta cần quan tâm đến yếu tố thời giá của dòng tiền.

Trong tài chính công thức tính NPV như sau:

Trong đó:

$$NPV = 0 = \sum_{t=1}^n \frac{B_t - C_t}{(1+r)^t}$$

Bt: Các khoản thu ở năm thứ t

Ct: Các khoản chi ở năm thứ t

n: Thời gian hoạt động của dự án

r: tỷ suất hiện tại hoá

Trên máy tính phần mềm E XCEL cú pháp chạy hàm NPV có dạng:

Cú pháp: = NPV(Rate, Value1, Value2, ..., ValueN)

Trong đó: Rate là lãi suất chiết khấu trong chiều dài chu kỳ.

Value1, Value2, ... là các số biểu hiện các khoản thu, chi, chúng phải cùng độ dài thời gian và xảy ra tại cuối mỗi kỳ thanh toán.

Thí dụ 5: Có 1 dự án trồng cây lâu năm có các dữ kiện thông tin về các khoản thu chi như sau: Chi phí ban đầu của dự án là: 119 triệu

Số tiền sản phẩm thu được năm thứ nhất là 50 triệu đồng, năm thứ 2 là 70 triệu đồng, năm thứ 3 là 30 triệu đồng. Thời gian hữu dụng của dự án: 3 năm. Giá trị tận thu vào năm cuối: 20 triệu đồng. Tỷ lệ lãi suất sử dụng tiền là 14%

Giải bài tập này bằng các công thức trong tài chính thể hiện trên bảng tính sau.

DVT: triệu đồng

Chi tiêu	Số tiền thu hoặc, chi	Thời gian Của tiền	hệ số hiện tại hoá (r=14%)	Giá trị hiện tại hoá của thu	Giá trị hiện tại hoá của chi
1. Vốn đầu tư ban đầu	119	hiện tại	1		119.000
2. Thu hàng năm	50	cuối năm thứ1	0.877	43.850	
	70	cuối năm thứ2	0.769	53.830	
	30	cuối năm thứ3	0.675	20.250	
3. Tận thu	20	cuối năm thứ3	0.675	13.500	
Tổng cộng				131.430	119.000
4. NPV	12.430				

Chạy hàm NPV trên EXCEL ta có kết quả như sau:

d) Hàm XNPV (Net Present Value) - Tổng giá trị thực tại hóa ròng

Đây là hàm tính giá trị thực tại hóa ròng của 1 dự án (hay còn gọi tổng lợi nhuận thực tại ròng của dự án hay vốn đầu tư) khi các dòng tiền xuất hiện không theo chu kỳ (vốn đầu tư theo năm, sản phẩm thu theo tháng, quý,...). Như vậy, XNPV cho phép chúng ta tính giá trị thực hiện tại cho 1 bảng thu chi tài chính mà không cần phải theo định kỳ.

Cú pháp: =XNPV(Rate,Values,Dates).

Trong đó:

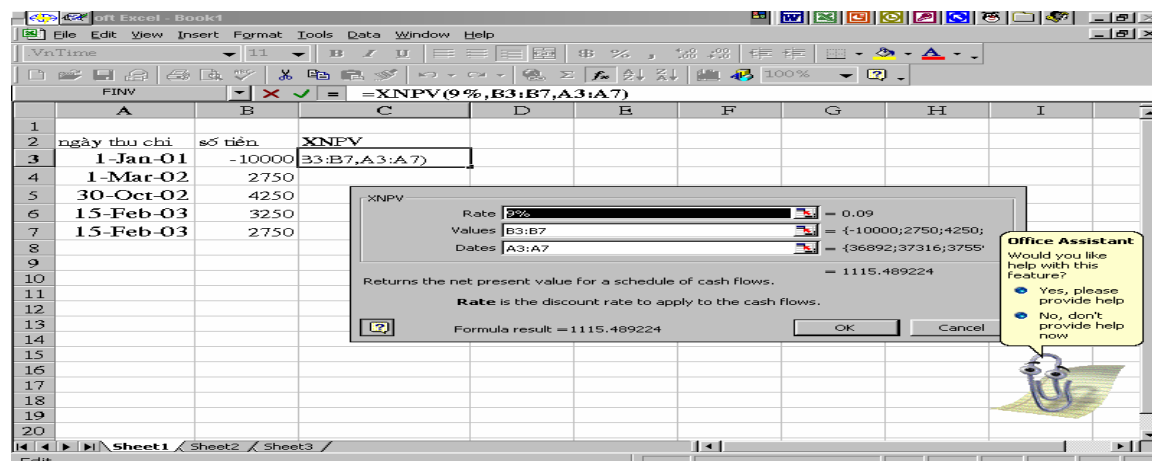
- Rate là suất chiết khấu tính được gán cho bảng thu chi tài chính.
- Values là chuỗi dòng tiền tệ tương ứng với một bảng thu chi theo ngày. Khoản đầu tiên là tùy chọn và tương ứng với giá trị hoặc chi phí khởi đầu của đầu tư. Các khoản thu chi sau được chiết khấu dựa trên cơ sở 365 ngày/ năm.
- Dates là bảng ngày thu chi tương ứng với 1 khoản thu, chi. Ngày thu chi đầu tiên chỉ định ngày bắt đầu bảng thu chi. Tất cả các ngày khác phải sau ngày này nhưng chúng có thể nằm theo bất cứ thứ tự nào.

Chú ý: Các số chỉ ngày được cắt bỏ phần thập phân.

Thí dụ: Giả sử đầu tư 1 khoản tiền là \$10000 vào ngày 1-Jan-2001 và sau đó thu được các giá trị hoàn vốn như sau:

- Ngày 1-Mar-2001 thu được là \$2750.
- Ngày 30-Oct-2002 thu được là \$4250.
- Ngày 15-Feb-2003 thu được là \$3250.
- Ngày 1-Apr-2003 thu được \$2750.

Giả sử lãi suất chiết khấu là 9%/ năm. Vậy giá trị hiện tại ròng của đầu tư này là bao nhiêu ?
Bài toán được chạy trên máy như sau:



e) Hàm IRR (Internal Rate of Return) - Tỷ suất nội hoàn vốn

Đây là hàm tính tỷ suất nội hoàn vốn (hay còn gọi là tỷ suất sinh lợi nội) của 1 chuỗi các thu chi tài chính từ một dự án đầu tư.

Tỷ suất sinh lợi nội là tỷ suất hiện tại hoá mà theo đó giá trị thực tại hóa ròng bằng 0 (hoặc chỉ số lợi ích trên chi phí bằng 1).

Từ công thức

$$NPV = 0 = \sum_{t=1}^n \frac{B_t - C_t}{(1+r)^t}$$

Trong đó:

Bt: Các khoản thu ở năm thứ t

Ct: Các khoản chi ở năm thứ t

n: Thời gian hoạt động của dự án

r: tỷ suất hiện tại hoá

Cụ thể, tính IRR theo công thức toán tài chính như sau:

$$IRR = r_l + \frac{(r_h - r_l) \cdot NPV(r_l)}{|NPV(r_l)| + |NPV(r_h)|}$$

Trong đó: r_l là tỷ suất hiện tại hoá thấp nhất

r_h là tỷ suất hiện tại hoá cao nhất

Có thể tính hàm IRR trên EXCEL dựa vào cú pháp sau:

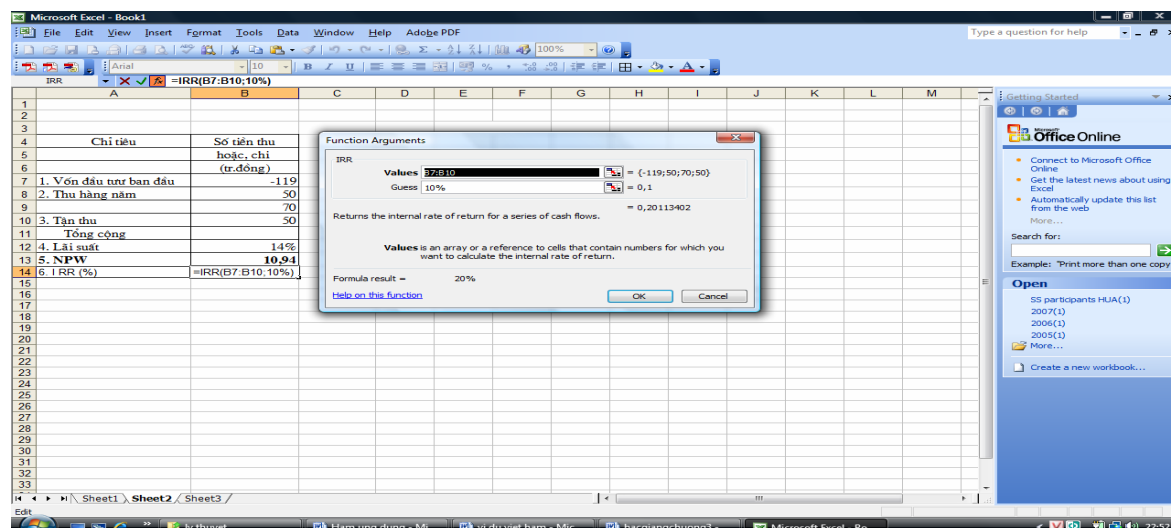
Cú pháp: = IRR(Values,Gueess)

Trong đó:

- Values là chuỗi tiền tệ tương ứng với 1 bảng thu chi tài chính. Values phải có ít nhất 1 giá trị dương (thu nhập) và 1 giá trị âm (chi phí) để tính suất nội tại hoàn vốn.

- Guees là kỳ vọng tỷ lệ tự hoàn vốn mong muốn, là 1 số mà ta dự đoán rằng nó gần với kết quả của hàm IRR. Trong nhiều trường hợp ta không cần cung cấp giá trị Guees, thì máy tính được mặc định bằng 10%.

Thí dụ: Lấy lại dữ liệu ở thí dụ 5, chạy hàm IRR trên EXCEL ta có kết quả như sau:



Bài tập thực hành: Giả sử anh H muốn kinh doanh khách sạn. Tiền đầu tư ban đầu là \$70000 và dự đoán thu nhập qua các năm lần lượt là \$12000; \$15000; \$18000; \$21000; \$26000. Yêu cầu tính tỷ suất nội hoàn của dự án sau 5 năm.

Như vậy bài toán này ta ngầm định Guess=10%, Ta có thể giải trên máy tính một cách dễ dàng.

f) Hàm MIRR (Mean Internal Rate of Return): Tính tỷ suất hoàn vốn của 1 dự án khi 2 dòng tiền âm và dương (hay dòng đầu tư và dòng lợi nhuận) có lãi suất khác nhau. Như vậy, MIRR sẽ trả về suất hoàn vốn nội tại cho 1 loạt các thu chi tài chính theo chu kỳ. Hàm MIRR xem xét cả giá đầu tư lẫn lãi suất nhận được trong tái đầu tư (reinvestment).

Theo công thức tính NPV ở trên chúng ta không thể tính chính xác tỷ suất nội hoàn vốn của một dự án mà chỉ có thể tính IRR bằng những ước lượng khoảng liên tiếp. Vì vậy, hàm MIRR được áp dụng trong trường hợp xem xét lãi suất tái đầu tư là phù hợp nhất.

- Công thức tính MIRR được biểu diễn bằng phương trình toán tài chính như sau:

$$MIRR = \left(\frac{-NPV(rrate, values[positive]) * (1 + rrate)^n}{NPV(frate, values[negative]) * (1 + frate)} \right)^{\frac{1}{n-1}} - 1$$

Trong đó: rrate là lãi suất từ tái đầu tư mang lại

frate là lãi suất đi vay

n là số kỳ ghép lãi của việc tái đầu tư

Nếu chạy hàm MIRR trên EXCEL thì khai báo cú pháp lệnh là:

Cú pháp: =MIRR(Values,Finance_rate,Reinvestment_rate)↵

Trong đó:

- Values là một dãy những giá trị biểu thị 1 loạt các thanh toán (giá trị âm) và thu nhập (giá trị dương) xảy ra trong những chu kỳ đều đặn. Values phải chứa ít nhất 1 giá trị dương và 1 giá trị âm để tính toán suất nội tại hoàn vốn.

- Finance_rate là lãi suất có NPV dương

- Reinvest_rate là lãi suất có NPV giá trị âm

Thí dụ: sử dụng số liệu ở bảng trên chạy hàm MIRR ta có kết quả như sau:

MIRR

Values: B2:B8 = {-1000;200;200;200}

Finance_rate: 8% = 0.08

Reinvest_rate: 5% = 0.05

= 0.052632665

Returns the internal rate of return for a series of periodic cash flows, considering both cost of investment and interest on reinvestment of cash.

Reinvest_rate is the interest rate you receive on the cash flows as you reinvest them.

Formula result = 0.052632665

OK Cancel

Năm (t)	Bt-Ct	r=5%	PV (5%)	R=8%	PV(8%)
1	-1000	1.05	-952.38	1.08	-925.93
2	200	1.1	181.82	1.17	170.94
3	200	1.16	172.41	1.26	158.73
4	200	1.22	163.93	1.36	147.06
5	200	1.28	156.25	1.47	136.05
6	200	1.34	149.25	1.59	125.79
7	200	1.41	141.84	1.71	116.96

Để kiểm tra tỷ lệ nội hoàn vốn của đầu tư trên chúng ta có thể tính lại NPV. Lấy lại thí dụ trên ta có bảng sau:

NPV

Rate: 5.26% = 0.0526

Value1: B2:B8 = {-1000;200;200;200}

Value2: number = 6.414603059

Returns the net present value of an investment based on a discount rate and a series of future payments (negative values) and income (positive values).

Value1: value1,value2,... are 1 to 29 payments and income, equally spaced in time and occurring at the end of each period.

Formula result = 6.414603059

OK Cancel

Năm (t)	Bt-Ct	r=5%	PV (5%)	R=8%	PV(8%)	R=5.26%
1	-1000	1.05	-952.38	1.08	-925.93	
2	200	1.1	181.82	1.17	170.94	
3	200	1.16	172.41	1.26	158.73	
4	200	1.22	163.93	1.36	147.06	
5	200	1.28	156.25	1.47	136.05	
6	200	1.34	149.25	1.59	125.79	
7	200	1.41	141.84	1.71	116.96	
NPV			\$14.42	(\$69.84)		

Bài tập thực hành: Một người đánh cá đã hoạt động được 5 năm. 5 năm trước anh ta mượn một số tiền là \$120000 với lãi suất 10%/ năm để mua 1 chiếc tàu. Việc đánh bắt cá của anh ta đạt được lợi nhuận là \$39000, \$30000, \$21000, \$37000 và \$46000. Trong những năm này anh ta tái đầu tư để mua sắm những thiết bị cần thiết cho việc đánh bắt cá và đã đạt được lợi nhuận là 12%/ năm. Hãy tính suất hồi vốn sau 3 năm, sau 5 năm.

g) Hàm XIRR (Internal rate of return): Tính lãi suất hoàn vốn của dự án khi các dòng tiền xuất hiện không theo chu kỳ (vốn đầu tư theo năm, sản phẩm thu theo tháng, quý,...). Như

vậy hàm XIRR sẽ trả về suất nội tại hoàn vốn cho 1 bảng thu chi tài chính mà không cần phải theo định kỳ.

Cú pháp: =XIRR(Value,Dates,Guess)↵

- Values là chuỗi dòng tiền tệ tương ứng với 1 bảng thu chi tài chính theo ngày. Khoản đầu tiên là tùy chọn và tương ứng với giá trị hoặc chi phí khởi đầu của đầu tư. Các khoản thu chi sau được chiết tính dựa trên cơ sở 365 ngày/ năm.

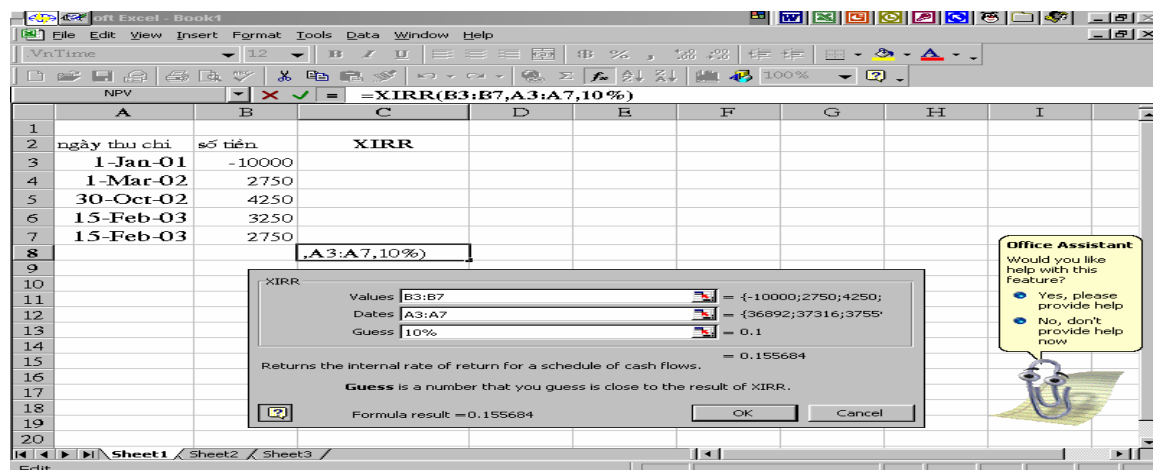
- Dates là bảng chuỗi thời gian (ngày) thu chi tương ứng với khoản thu chi. Ngày thu chi đầu tiên chỉ định ngày bắt đầu bảng thu chi. Tất cả các ngày khác phải sau ngày này nhưng chúng có thể nằm theo bất kỳ thứ tự nào.

- Guess là kỳ vọng tỷ suất nội hoàn vốn mong muốn (dự đoán rằng nó gần với kết quả của hàm XIRR). Trong nhiều trường hợp ta không cần cung cấp giá trị ước lượng để tính toán hàm XIRR, máy tự mặc nhận Guess = 10%.

Lưu ý: Các chỉ số chỉ ngày (số thứ tự) được máy mặc định cắt bỏ phần thập phân.

Hàm XIRR luôn mặc nhận dòng tiền có ít nhất 1 giá trị dương (thu) và 1 giá trị âm (chi).

Thí dụ: Lấy lại thí dụ trong phần tính XNPV ta tính hàm XIRR như sau:



h) Hàm PMT (Payment): Tính lượng tiền cho từng kỳ khi biết giá trị hiện tại hoặc tương lai của dòng tiền (tính mua trả góp, tiết kiệm). Như vậy, hàm PMT là lượng tiền trả định kỳ cho 1 khoản chi dựa trên số tiền phải trả cố định và lãi suất không thay đổi.

Cú pháp: =PMT(Rate,Nper,Pv,Fv,Type)↵

Trong đó:

- Rate là lãi suất/ kỳ.
- Nper là số lần trả tiền trong 1 năm.
- Pv là giá trị hiện tại.
- Fv là giá trị tương lai. Nếu không ghi Fv thì nó được coi bằng 0.
- Type nhận 2 giá trị: = 0 (trả cuối chu kỳ); = 1 (trả đầu kỳ). Máy mặc nhận = 0

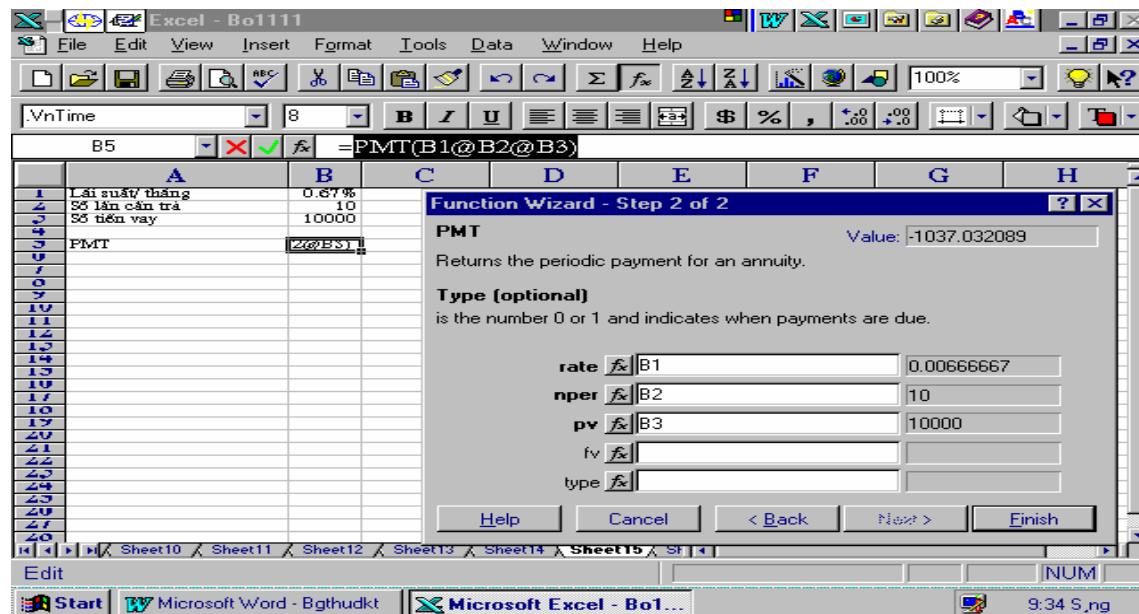
Chú ý: Giá trị thanh toán được trả về bởi PMT bao gồm gốc và lãi nhưng không có tính thuế hoặc những chi phí khác. Để tìm tổng số tiền thanh toán trong suốt 1 năm ta nhân kết quả của hàm PMT với Nper.

Thí dụ 7: Giả sử ngày 1/1/99 ta vay 1 khoản tiền là \$10000, với lãi là 8%/ năm. Hàng tháng ta phải trả 1 khoản tiền nhất định (1 phần gốc và lãi) và phải trả trong vòng 10 tháng (10 lần). Vậy khoản tiền phải trả mỗi lần là bao nhiêu.

Sử dụng hàm PMT trong EXCEL để giải bài toán.

Bài tập thực hành: Anh A mượn của chị T \$5000 với lãi suất 12%/ năm. Anh A hàng tháng phải trả cho chị T một khoản tiền nhất định trong vòng 5 tháng. Chị T và anh A tính mãi không ra, họ nhờ bạn tính hộ ?

Giả sử ta muốn tiết kiệm \$50000 trong 18 năm bằng cách để dành 1 khoản tiền cố định mỗi tháng. Giả sử ta có khả năng kiếm được 6% lãi trên số tiền cần tiết kiệm. Vậy số tiền tiết kiệm hàng tháng cụ thể là bao nhiêu USD ?



1. 3. Một số hàm thông dụng khác

a) Hàm VLOOKUP hoặc HLOOKUP

* **Hàm VLOOKUP** (Vertical Look Up): Là hàm tìm kiếm giá trị nào đó theo cột, nó định vị giá trị của ô theo bảng số liệu cho trước, vì vậy nó có nhiều ứng dụng trong kinh tế như tính thuế, tính lương theo các mức giá trị nhất định.

Cú pháp:

= VLOOKUP(Lookup_value,Table_array,Col_index_num,Range_lookup)

Trong đó:

- **Lookup_value** (ký hiệu là X): Là giá trị cần dò tìm, khi tìm đem so sánh với giá trị cột bên trái của bảng, nếu có thì lệch qua bên phải đến cột tham chiếu để lấy giá trị trong ô ở đó ứng với vị trí của X.
- **Table_array** (bảng): Bảng là 1 khối các ô, thông thường bảng gồm nhiều hàng và nhiều cột. Cột bên trái luôn luôn chứa các địa chỉ để dò tìm, các cột khác chứa các giá trị tương ứng để tham chiếu.
- **Col_index_num** (cột tham chiếu): Cột tham chiếu là thứ tự của cột (tính từ phía trái sang phía phải của bảng table_array), cột đầu tiên của bảng là cột 1.
- **Range_lookup** (cách dò): Cách dò là số 0 hoặc số 1. Ngầm định là số 1.

+ Cách dò là số 1:

- Danh sách ở cột bên trái của bảng phải sắp xếp theo thứ tự tăng dần.
- Nếu giá trị dò X nhỏ hơn phần tử đầu tiên trong danh sách hàm cho giá trị là #N/A (Not Available: bất khả thi).
- Nếu giá trị dò X đúng khớp với 1 phần tử trong danh sách (không phân biệt chữ hoa hay thường nếu là chuỗi), đương nhiên tìm thấy ở tại phần đó và cho trị của ô nằm trong cột tham chiếu cùng hàng với phần tử này.

+ Nếu cách dò là 0:

- Danh sách ở cột bên trái của bảng không cần phải xếp theo thứ tự.
- Nếu giá trị dò X không đúng khớp với bất kỳ phần tử nào trong danh sách (không phân biệt chữ hoa hay thường nếu là chuỗi), hàm cho trị là #N/A (Not Available: bất khả thi).

- Chỉ khi nào trị dò X đúng khớp với 1 phần tử trong danh sách (không phân biệt chữ hoa hay thường nếu là chuỗi) mới cho trị là trị của ô nằm trong cột tham chiếu cùng hàng với phần tử này.

Ví dụ: Tính lương cho cán bộ trong 1 tháng theo số ngày làm việc với 4 đơn giá lương khác nhau là: Loại A: 200 ngàn đồng/ ngày; Loại B: 150 ngàn đồng/ ngày; Loại C: 100 ngàn đồng/ ngày và loại D là 50 ngàn đồng/ ngày.

Tổng tiền lương 1 người 1 tháng = Số ngày làm việc * đơn giá lương ngày

Giải bài tập này trên EXCEL như sau :

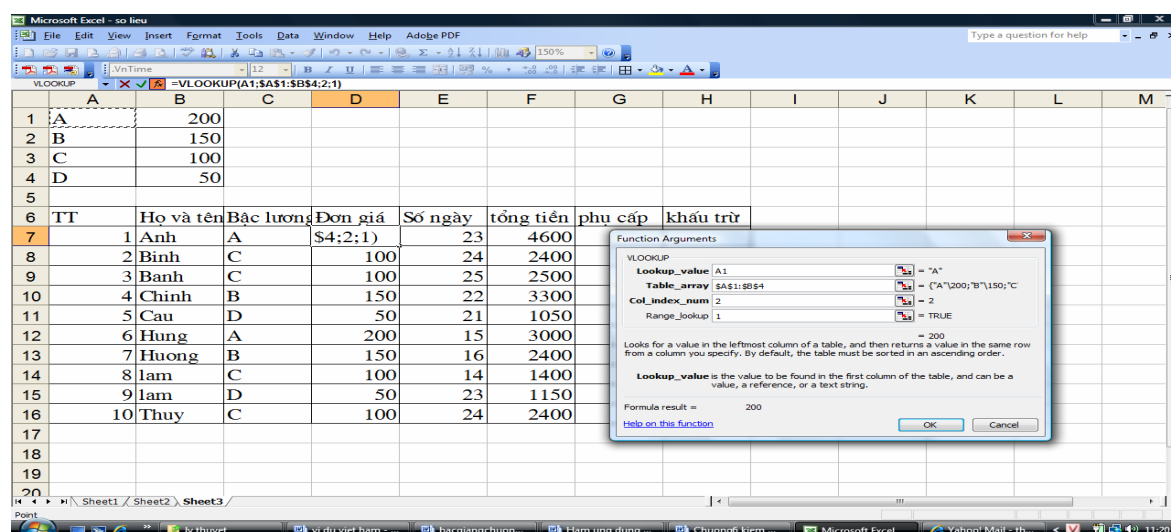
- Lập bảng table_array: trong thí dụ này gồm 2 cột và 4 hàng (A1:B4)

- Thiết kế bảng lương của cán bộ trong từng tháng (trong ví dụ này bảng lương có thể lập có định gồm nhiều cột và hàng. Chẳng hạn có cột thứ tự, Họ và tên, bậc lương, số ngày làm việc, đơn giá lương, tổng tiền, các khoản phụ cấp, khấu trừ.....

- Xác định cột cần tìm giá trị: trong ví dụ này tìm cột đơn giá lương (D7:D16)

- Vào Insert/Vlookup và khai báo các thông tin cần thiết.

Tất cả các bước này có thể tham khảo trên hình sau:



Chú ý: Phải lấy địa chỉ tuyệt đối của bảng cần tham chiếu dò tìm (\$A\$1:\$B\$4).

* **Hàm HLOOKUP** (Horizontal Look Up): Là hàm tìm kiếm, nó định vị giá trị của ô theo bảng số liệu cho trước theo hàng, vì vậy nó có nhiều ứng dụng trong kinh tế như hàm Vlookup trong tính thuế, tính lương, v.v...

Cú pháp:

= **HLOOKUP**(Lookup_value, Table_array, Row_index_num, Range_lookup)

Trong đó:

- Lookup_value (ký hiệu là X): Là giá trị cần dò tìm, khi tìm đem so sánh với hàng trên cùng của bảng, nếu có thì ứng với hàng dưới đến dòng tham chiếu để lấy giá trị trong ô ứng với vị trí của X.

- Table_array (bảng): Bảng là 1 khối các ô, thông thường bảng gồm nhiều hàng và nhiều cột. Dòng trên cùng luôn luôn chứa các địa chỉ để dò tìm, các dòng khác chứa các giá trị tương ứng để tham chiếu.

- Row_index_num (dòng của bảng Table_array): Dòng tham chiếu là thứ tự của dòng (tính từ trên xuống của bảng), dòng đầu tiên của bảng là dòng 1.

- Range_lookup (cách dò): Cách dò là số 0 hoặc số 1. Ngầm định là số 1.

* **Cách dò là số 1:**

- Danh sách ở dòng trên cùng của bảng phải sắp xếp theo thứ tự tăng dần.

- Nếu trị dò X nhỏ hơn phần tử đầu tiên trong danh sách hàm cho trị là #N/A (Not A available: bất khả thi).

- Nếu trị dò X đúng khớp với 1 phần tử trong danh sách (không phân biệt chữ hoa hay thường nếu là chuỗi), đương nhiên tìm thấy ở tại phần đó và cho trị của ô nằm trong dòng tham chiếu cùng cột với phần tử này.

*** Nếu cách dò là 0:**

- Danh sách ở dòng trên cùng của bảng không cần phải xếp theo thứ tự.
- Nếu trị dò X không đúng khớp với bất kỳ phần tử nào trong danh sách (không phân biệt chữ hoa hay thường nếu là chuỗi), hàm cho trị là #N/A (Not Available: bất khả thi).
- Chỉ khi nào trị dò X đúng khớp với 1 phần tử trong danh sách (không phân biệt chữ hoa hay thường nếu là chuỗi) mới cho trị là trị của ô nằm trong hàng tham chiếu cùng cột với phần tử này.

Như vậy, mọi nguyên tắc hoạt động của hàm này giống như hàm VLOOKUP chỉ khác:

- Hàm VLOOKUP dò tìm ở cột bên trái, tham chiếu số liệu ở các cột bên phải.
- Hàm HLOOKUP dò tìm ở hàng trên cùng, tham chiếu số liệu ở các hàng phía dưới.

Cũng ví dụ trên ta xử lý như sau:

The screenshot shows an Excel spreadsheet with the following data in columns A-D (rows 5-14):

TT	Họ và tên	Bậc lương	Đơn giá
1	Anh	A	\$2;2;1
2	Binh	C	100
3	Banh	C	100
4	Chinh	B	150
5	Cau	D	50
6	Hung	A	200
7	Huong	B	150
8	lam	C	100
9	lam	D	50
10	Thuy	C	100

The HLOOKUP function is being used in cell C5 with the formula: `=HLOOKUP(C5,$A$1:$D$2,2,1)`. The dialog box shows the following arguments:

- Lookup_value: 55
- Table_array: \$A\$1:\$D\$2
- Row_index_num: 2
- Range_lookup: 1

The formula result is 200.

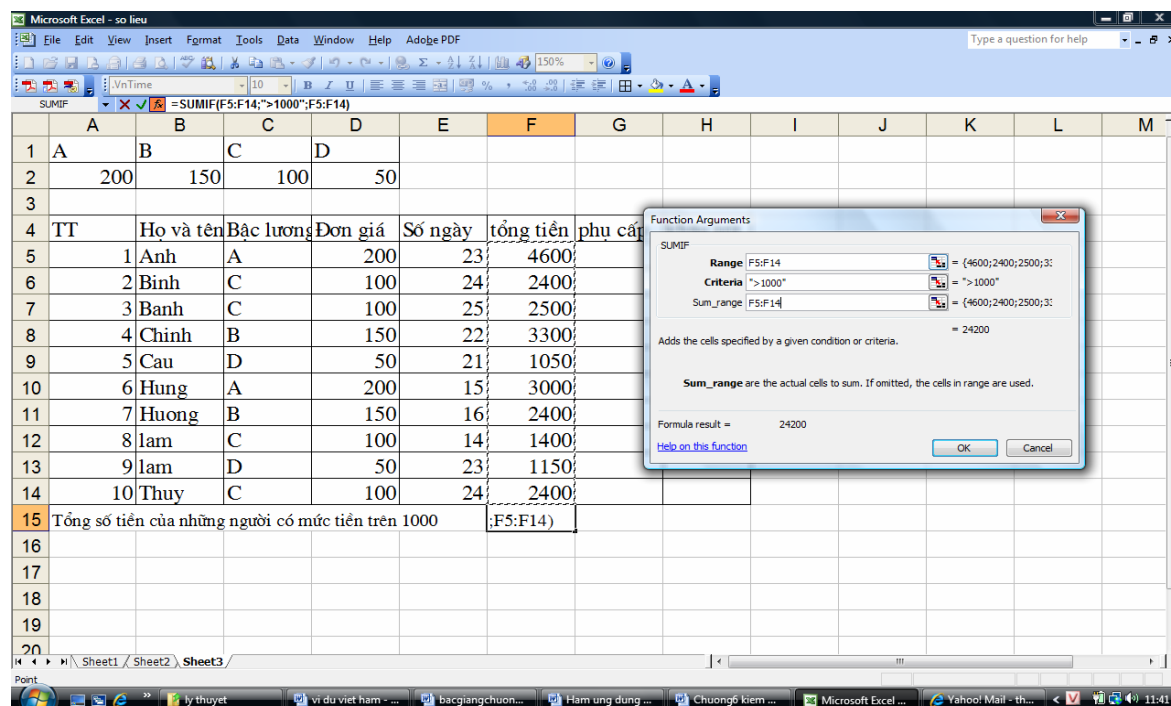
Chú ý: Phải lấy địa chỉ tuyệt đối của bảng cần tham chiếu dò tìm (\$A\$1:\$D\$2).

b) Hàm SUMIF: Tính tổng của miền dữ liệu thỏa mãn điều kiện nào đó.

Cú pháp: `=SUMIF(Range,Criteria,Sum_range)`

- Range: Miền dữ liệu cần phải thỏa mãn điều kiện (Criteria)
- Criteria: Điều kiện đặt ra trong việc tính tổng.
- Sum_range: Miền cần tính tổng các số thỏa mãn điều kiện.

Thí dụ: Hãy tính tổng tiền lương của những người có mức lương trên 1000 đồng



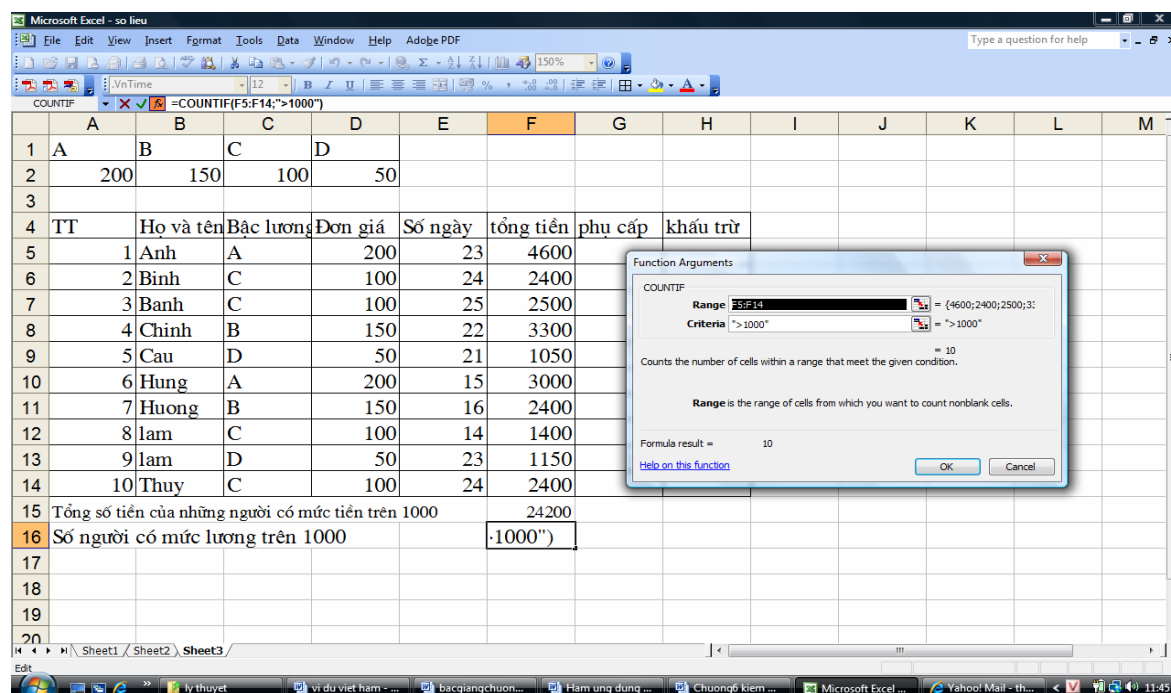
c) Hàm COUNTIF: Đếm miền dữ liệu thỏa mãn điều kiện nào đó.

Cú pháp: =COUNTIF(Range,Criteria)↵

- Range: Phạm vi miền dữ liệu cần đếm.

- Criteria: Điều kiện đặt ra đối với miền dữ liệu khi đếm.

Thí dụ: Đếm số người có mức lương trên 1000 đồng



d) Các hàm thống kê

Các hàm thống kê được sử dụng rất phổ biến trong xử lý và tổng hợp các số liệu thí nghiệm.

Thí dụ: Quan sát năng suất hái chè búp tươi của 4 nhóm công nhân thuộc Công ty chè Phú Đa trong 1 ngày như sau.

Thứ tự	Năng suất lao động (kg/ngày)			
	Nhóm 1	Nhóm 2	Nhóm 3	Nhóm 4

1	50	52	46	40
2	48	46	47	45
3	45	50	56	46
4	60	65	54	50
5	70	78	45	55
6	62	61	51	60
7	55	58	58	47
8	62	70	47	67
9	58	67	43	70
10	53	65	62	80

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1																
2	TT	Hàm	Nhóm 1	Nhóm 2	Nhóm 3	Nhóm 4	Số ngày	xlf1								
3	1		50	52	46	40	4	200								
4	2		48	46	47	45	5	240								
5	3		45	50	56	46	6	270								
6	4		60	65	54	50	7	420								
7	5		70	78	45	55	3	210								
8	6		62	61	51	60	4	248								
9	7		55	58	58	47	5	275								
10	8		62	70	47	67	6	372								
11	9		58	67	43	70	3	174								
12	10		53	65	62	80	4	212								
13	Tổng số quan sát	= COUNT(...)	10,00	10,00	10,00	10,00	10,00									
14	Tổng cộng	= SUM(.....)	573,00	622,00	519,00	570,00	57,00	2621								
15	Tổng các tích	= SUMPRODUCT(...)						2621								
16	Giá trị cao nhất	= MAX(.....)	70,00	78,00	62,00	80,00	7,00									
17	Giá trị thấp nhất	= MIN(.....)	45,00	46,00	43,00	40,00	3,00									
18	Trung bình	= AVERAGE(.....)	56,30	61,20	50,90	56,00	4,70									
19	Phương sai mẫu	= VAR(.....)	57,57	97,07	40,10	167,11	1,79									
20	Phương sai tổng thể	= VARP(.....)	51,81	87,36	36,09	150,40	1,61									
21	Độ lệch chuẩn mẫu	= STDEV(.....)	7,59	9,85	6,33	12,93	1,34									
22	Độ lệch chuẩn tổng thể	= STDEVP(.....)	7,20	9,35	6,01	12,26	1,27									
23	Độ nhọn	= KURT(.....)	-0,37	-0,55	-1,00	-0,56	-0,85									
24	Độ bất đối xứng	= SKEW(.....)	0,22	0,00	0,52	0,68	0,33									
25	Số trung vị	= MEDIAN(.....)	56,50	63,00	49,00	52,50	4,50									
26	Mốt	= MOD(.....)														
27	Hệ số chặn	= INTERCEPT(.....)	164,22													
28	Hệ số góc	= SLOP(.....)														
29	Hệ số tương quan	= CORREL(.....)														
30	Hiệp phương sai	= COVAR(.....)														

2. Các lệnh ứng dụng trong EXCEL

2.1. Cụm lệnh TOOLS/DATA ANALYSIS - Cụm lệnh phân tích số liệu

Ngoài việc dùng một số hàm ở trong Category Statistical (khi khai thác các hàm Function) việc phân tích số liệu chủ yếu thực hiện qua thư mục Data Analysis của menu Tools. Nếu trên menu Tools không thấy mục này thì mở ở mục Add-In .

Một số thuật ngữ thường dùng trong phân tích số liệu của thư mục Data Analysis là:

Input range (Miền vào): Tùy theo từng mục mà chỉ ra ô đầu ...ô cuối

Thí dụ: Biến phụ thuộc Y: D1:D20

Các biến độc lập: A1:C20

Label (nhãn): Chủ yếu là tên biến tên các mức độ của nhân tố. Nếu miền vào có nhãn ở đầu các cột (hay đầu các dòng) số liệu thì phải nháy vào ô Label.

Confidence level (Mức tin cậy): Thông thường trong máy tính đã mặc định mức tin cậy $P=0,95\%$, nhưng đôi khi trong một số phân tích số liệu người ta sử dụng mức sai số cho phép $\alpha = 1 - P = 0,05$ hay 5% . Ta có thể chọn mức tin cậy khác tùy mục đích như $P = 0,99$ hay $P = 0,90$.

Output range (Miền ra): Có thể lấy miền ra là một trang mới (new sheet), một tệp mới (new workbook), nhưng người ta thường chọn miền ra là một khu vực còn trống trên trang số liệu đang tính toán và chỉ cần nêu ô đầu tiên mà không cần nêu cả miền.

Option (tùy chọn): Một số phần ta có thể chọn (Nháy đen hoặc đánh dấu x) hoặc bỏ qua (để trống).

Sau đây là một số ứng dụng của cụm lệnh này trong phân tích số liệu.

a) Lệnh TOOLS/ DATA ANALYSIS/ ANOVA – Phân tích phương sai

* Phân tích phương sai một yếu tố - Single Factor

Phân tích phương sai một yếu tố là phân tích ảnh hưởng của một yếu tố nguyên nhân (thường là yếu tố định tính) đến một yếu tố kết quả (thường là yếu tố định lượng) đang nghiên cứu.

Giả sử chúng ta cần so sánh số trung bình của k tổng thể độc lập. Người ta lấy k mẫu có số quan sát là $n_1, n_2, \dots, n_k$; tuân theo phân phối chuẩn. Trung bình của các tổng thể được ký hiệu là $\mu_1, \mu_2, \dots, \mu_k$ thì mô hình phân tích phương sai một yếu tố ảnh hưởng được mô tả dưới dạng kiểm định giả thuyết có dạng như sau:

Ho: $\mu_1 = \mu_2 = \dots = \mu_k$

H1: Tồn tại ít nhất 1 cặp có $\mu_i \neq \mu_j$ với $i \neq j; i = 1, 2, \dots, k$ và $j = 1, 2, \dots, k$

Để kiểm định ta đưa ra 2 giả thiết sau:

1) Mỗi mẫu tuân theo phân phối chuẩn $N(\mu, \sigma^2)$

2) Ta lấy k mẫu độc lập từ k tổng thể. Mỗi mẫu được quan sát n_j lần.

Về trình tự phân tích phương sai được chia thành 4 bước.

Bước 1: Tính các trung bình mẫu và trung bình chung của k mẫu

Bước 2: Tính các tổng độ lệch bình phương

Bước 3: Tính các phương sai (phương sai của nội bộ nhóm và phương sai giữa các nhóm)

Bước 4: Kiểm định giả thuyết

- Tính tiêu chuẩn kiểm định F (F thực nghiệm)

$$F = \frac{MSB}{MSW}$$

Trong đó: MSB là phương sai giữa các nhóm
MSW là phương sai trong nội bộ nhóm

- Tìm F lý thuyết (F tiêu chuẩn = $F(\alpha, k-1, n-k)$): là giá trị giới hạn tra từ bảng phân phối F với k-1 bậc tự do của phương sai ở tử số và n-k bậc tự do của phương sai ở mẫu số với mức ý nghĩa α . F lý thuyết có thể tra qua hàm FINV($\alpha, k-1, n-k$) trong EXCEL.

- Nếu F thực nghiệm > F lý thuyết, bác bỏ Ho, nghĩa là các số trung bình của k tổng thể không bằng nhau.

Bảng phân tích phương sai 1 yếu tố có thể tóm tắt như sau:

Bảng gốc bằng tiếng Anh

Source of variation	Sum of squares (SS)	Degree of freedom (df)	Mean squares (MS)	F- ratio
Between- groups	SSB	(k-1)	MSB=SSB/k-1	F = $\frac{MSB}{MSW}$
Within-groups	SSW	(n-k)	MSW = SSW/n-k	
Total	SST	(n-1)		

Bảng phân tích phương sai tổng quát dịch ra tiếng việt – ANOVA

Nguồn biến động	Tổng độ lệch bình phương (SS)	Bậc tự do (df)	Phương sai (MS)	F- thực nghiệm
Giữa các mẫu	SSB	(k-1)	MSB	F = $\frac{MSB}{MSW}$
Trong nội bộ các mẫu	SSW	(n-k)	MSW	
Tổng số	SST	(n-1)		

Thí dụ 1: Có tài liệu về cách cho điểm môn lý thuyết thống kê của 3 giáo sư như sau (điểm tối đa là 100). Hãy cho biết cách chấm điểm của 3 giáo sư có sai khác nhau không?

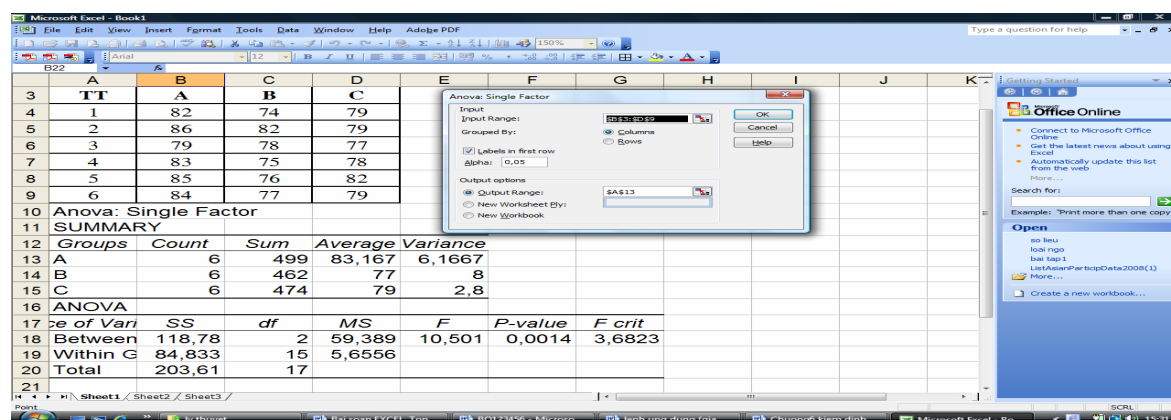
TT	A	B	C
1	82	74	79
2	86	82	79

3	79	78	77
4	83	75	78
5	85	76	82
6	84	77	79

Sử dụng phần mềm EXCEL giải bài toán này qua các bước sau:

- Thiết kế dữ liệu trên bảng tính EXCEL(thường sắp theo cột – bảng không phân tổ, chỉ liệt kê các quan sát)
- Chọn Tools/Data analysis/ANOVA/Single Factor
- Khai báo Input range; Label ; Alfa ; Output range
- OK

Các bước thực hiện trên EXCEL được thể hiện qua hình dưới đây.



* Phân tích phương sai 2 yếu tố không lặp- Two Factors without Replication

Phân tích phương sai 2 yếu tố nhằm xem xét cùng một lúc hai yếu tố nguyên nhân (dưới dạng dữ liệu định tính) ảnh hưởng đến yếu tố kết quả (dưới dạng dữ liệu định lượng) đang nghiên cứu.

Thí dụ: Nghiên cứu ảnh hưởng của loại chất đốt và loại lò sấy đến tỷ lệ vải loại 1 sấy khô.

Phân tích phương sai 2 hai yếu tố giúp chúng ta đưa thêm yếu tố nguyên nhân vào phân tích làm cho kết quả nghiên cứu càng có giá trị. Song, đòi hỏi chúng ta phải sắp xếp dữ liệu thành bảng hai chiều.

Giả sử ta nghiên cứu ảnh hưởng của 2 yếu tố nguyên nhân định tính đến một yếu tố kết quả định lượng nào đó. Ta lấy mẫu không lặp lại, sau đó các đơn vị mẫu của yếu tố nguyên nhân thứ nhất sắp xếp thành K nhóm (cột), các đơn vị mẫu của yếu tố nguyên nhân thứ hai sắp xếp thành H khối (hàng). Như vậy, ta sẽ có bảng kết hợp 2 yếu tố nguyên nhân gồm K cột và H hàng và (KxH) ô dữ liệu. Tổng số mẫu quan sát sẽ là $n = (K \times H)$. Dạng tổng quát của bảng này như sau:

Bảng sắp xếp các mẫu quan sát của phân tích phương sai 2 yếu tố không lặp

Hàng (khối)	Cột (nhóm)			
	1	2	...	K
1	X_{11}	X_{21}	X_{31}	X_{K1}
2	X_{12}	X_{22}	X_{32}	X_{K2}
...	...	...	...	...
...	...	...	...	...
H	X_{1H}	X_{2H}	X_{3H}	X_{KH}

Mô hình phân tích phương sai hai yếu tố ảnh hưởng được mô tả dưới dạng kiểm định giả thuyết bao gồm 2 phần :

(1) Kiểm định giả thuyết H_0 : cho số trung bình của K tổng thể, tương ứng với K nhóm mẫu (k cột) là bằng nhau;

(2) Kiểm định giả thuyết H_0 : cho số trung bình của H tổng thể, tương ứng với H khối mẫu (hàng) là bằng nhau;

Để kiểm định ta đưa ra 2 giả thiết sau:

1) Mỗi mẫu tuân theo phân phối chuẩn $N(\mu, \sigma^2)$

2) Ta lấy K mẫu độc lập từ K tổng thể, H mẫu độc lập từ H tổng thể. Mỗi mẫu được quan sát 1 lần không lặp.

Phân tích phương sai 2 yếu tố không lặp cũng tuân theo 4 bước như phân tích phương sai 1 yếu tố. Riêng ở bước 4 - Kiểm định giả thuyết cần phải tính tiêu chuẩn kiểm định riêng cho từng yếu tố nguyên nhân. Cụ thể:

+ Tính tiêu chuẩn kiểm định F (F thực nghiệm)

$$F_1 = \frac{MSK}{MSE}$$

Trong đó: MSK là phương sai giữa các nhóm (cột)

MSE là phương sai phần dư

F_1 dùng kiểm định cho yếu tố nguyên nhân thứ nhất

$$F_2 = \frac{MSH}{MSE}$$

Trong đó: MSH là phương sai giữa các khối (hàng)

MSE là phương sai phần dư

F_2 dùng kiểm định cho yếu tố nguyên nhân thứ hai

+ Tìm F lý thuyết cho 2 yếu tố nguyên nhân;

- Yếu tố nguyên nhân thứ nhất: (F tiêu chuẩn = $F(k-1; (k-1)(h-1), \alpha)$) là giá trị giới hạn tra từ bảng phân phối F với k-1 bậc tự do của phương sai ở tử số và (k-1)(h-1) bậc tự do của phương sai ở mẫu số với mức ý nghĩa α . F lý thuyết có thể tra qua hàm FINV($\alpha, k-1, (k-1)(h-1)$) trong EXCEL.

- Yếu tố nguyên nhân thứ hai: (F tiêu chuẩn = $F(h-1, (k-1)(h-1), \alpha)$) là giá trị giới hạn tra từ bảng phân phối F với (h-1) bậc tự do của phương sai ở tử số và (k-1)(h-1) bậc tự do của phương sai ở mẫu số với mức ý nghĩa α . F lý thuyết có thể tra qua hàm :

FINV($\alpha, h-1, (k-1)(h-1)$) trong EXCEL.

- Nếu F_1 thực nghiệm > F_1 lý thuyết thì bác bỏ H_0 cho rằng các số trung bình của k tổng thể nhóm (cột) bằng nhau.

- Nếu F_2 thực nghiệm > F_2 lý thuyết thì bác bỏ H_0 cho rằng các số trung bình của k tổng thể khối (hàng) bằng nhau.

Bảng phân tích phương sai 2 yếu tố khi sử dụng máy tính (phần mềm EXCEL hoặc SPSS) tóm tắt như sau:

Bảng gốc bằng tiếng Anh

Source of variation	Sum of squares (SS)	Degree of freedom (df)	Mean squares (MS)	F-ratio
Rows	SSH	(h-1)	MSH	F_1
Columns	SSK	(k-1)	MSK	F_2
Error	SSE	(k-1)(h-1)	MSE	
Total	SST	(n-1)		

Bảng phân tích phương sai tổng quát dịch ra tiếng Việt – ANOVA

Nguồn biến động	Tổng độ lệch bình phương (SS)	Bậc tự do (df)	Phương sai (MS)	F- Tỷ số
Giữa các hàng	SSH	(h-1)	MSH	F_1
Giữa các cột	SSK	(k-1)	MSK	F_2
Phần dư	SSE	(k-1)(h-1)	MSE	
Tổng số	SST	(n-1)		

Thí dụ 2:

Có tài liệu về giá bán đậu tương của các tỉnh qua 2 năm như sau (đồng/kg)

Tỉnh	2003	2004
Sơn La	4440	4247,7
Hà Tây	4850	4294,3
Đắc Lắc	4400	4284,3
Đồng nai	4500	4314,3

Yêu cầu: Sử dụng kết quả phân tích phương sai so sánh giá bán đậu tương qua 2 năm và giữa 4 tỉnh?

Giải: Sử dụng phân tích phương sai (ANOVA) 2 yếu tố lấy mẫu không lặp trong EXCEL như sau:

- Thiết kế dữ liệu thành bảng hai chiều trên EXCEL
- Chọn Tools/Data analysis/ANOVA/Two –Factors without Replication
- Khai báo **Input range; Label ; Alfa ; Output range**
- OK

Kết quả phân tích phương sai hai yếu tố không lặp được thể hiện qua 2 hình dưới đây.

Anova: Two-Factor Without Replication

Input Range: \$A\$3:\$C\$7

Labels: ☒ Labels

Alpha: 0,05

Output options:

- ☒ Output Range: \$A\$9
- ☐ New Worksheet Ply:
- ☐ New Workbook

Summary Table:

	Count	Sum	Average	Variance
Sơn La	2	8687,7	4343,85	18489,64
Hà Tây	2	9144,3	4572,15	154401,2
Đắc Lắc	2	8684,3	4342,15	6693,245
Đồng nai	2	8814,3	4407,15	17242,25
2003	4	18190	4547,5	42358,33
2004	4	17140,6	4285,15	778,89

Anova: Single Factor

Input Range: \$B\$3:\$D\$9

Labels in first row: ☒ Labels in first row

Alpha: 0,05

Output options:

- ☒ Output Range: \$A\$13
- ☐ New Worksheet Ply:
- ☐ New Workbook

Summary Table:

Groups	Count	Sum	Average	Variance
A	6	499	83,167	6,1667
B	6	462	77	8
C	6	474	79	2,8

ANOVA Table:

	SS	df	MS	F	P-value	F crit
Between Groups	118,78	2	59,389	10,501	0,0014	3,6823
Within Groups	84,833	15	5,6556			
Total	203,61	17				

* Phân tích phương sai 2 yếu tố có lặp – Two Factors with Replication

Phân tích phương sai 2 yếu tố có lập cũng giống như trường hợp phân tích phương sai hai yếu tố không lập, chỉ khác là mỗi thí nghiệm (hay quan sát) thường được lặp lại nhiều lần (2-3 lần).

Bảng sắp xếp các mẫu quan sát của phân tích phương sai 2 yếu tố có lập có dạng như sau:

Hàng (khối)	Cột (nhóm)			
	1	2	...	K
1	X_{11}	X_{21}	X_{31}	X_{K1}
	X_{11}	X_{21}	X_{31}	X_{K1}
2	X_{12}	X_{22}	X_{32}	X_{K2}
	X_{12}	X_{22}	X_{32}	X_{K2}
...	...	...	...	...
...	...	...	...	...
H	X_{1H}	X_{2H}	X_{3H}	X_{KH}
	X_{1H}	X_{2H}	X_{3H}	X_{KH}

Theo thiết kế của bảng này một yếu tố (A) xếp ở hàng (h hàng), một yếu tố (B) sắp theo cột (k cột), mỗi giá trị theo hàng và cột được lặp lại 1 lần (1 số lần lặp lại). Vì vậy bảng phân tích phương sai 2 yếu tố khi sử dụng máy tính (phần mềm EXCEL hoặc SPSS) tóm tắt như sau:

Bảng gốc bằng tiếng Anh

Source of variation	Sum of squares (SS)	Degree of freedom (df)	Mean squares (MS)	F- ratio	F tra bảng
Sample (A) Rows	SSH	(h-1)	MSH	F_1	$F_1(0.05, h-1, kh(l-1))$
Columns (B)	SSK	(k-1)	MSK	F_2	$F_2(0.05, k-1, kh(l-1))$
Interaction(AB)	SSHK	(k-1)(h-1)	MSHK	F_3	$F_3(0.05, (h-1)(k-1), kh(l-1))$
Within	SSE	kh(l-1)	MSE		
Total	SST	(n-1)			

Bảng phân tích phương sai tổng quát dịch ra tiếng việt – ANOVA

Nguồn biến động	Tổng độ lệch bình phương (SS)	Bậc tự do (df)	Phương sai (MS)	F Thực nghiệm	F tra Bảng
Giữa các hàng (A)	SSH	(h-1)	MSH	F_1	$F_1(0.05, h-1, kh(l-1))$
Giữa các cột (B)	SSK	(k-1)	MSK	F_2	$F_2(0.05, k-1, kh(l-1))$
Tương tác hàng và cột (AB)	SSHK	(k-1)(h-1)	MSHK	F_3	$F_3(0.05, (h-1)(k-1), kh(l-1))$
Sai số ngẫu nhiên	SSE	kh(l-1)	MSE		
Tổng số	SST	(n-1)			

Lấy lại thí dụ 2 trên đây, mỗi tỉnh ở mỗi năm lặp lại 2 giá trị tính về giá bán đậu tương của các tỉnh như sau (đồng/kg).

Tỉnh	2003	2004
Sơn La	4440	4247,7
	4530	4330,0
Hà Tây	4850	4294,3
	4789	4321,8
Đắc Lắc	4400	4284,3
	4500	4389,6
Đồng nai	4500	4314,3
	4450	4400,0

Yêu cầu: Sử dụng kết quả phân tích phương sai so sánh giá bán đậu tương qua 2 năm và giữa 4 tỉnh?

Giải: Sử dụng phân tích phương sai (ANOVA) 2 yếu tố lấy mẫu có lập trong EXCEL như sau:

- Thiết kế dữ liệu thành bảng hai chiều trên EXCEL
- Chọn Tools/Data analysis/ANOVA/Two-Factors with Replication
- Khai báo **Input range**; **Rows per Sample**; **Alfa** ;
- Output range**.
- OK

Anova: Two-Factor With Replication

Input Range: \$A\$2:\$C\$10

Rows per sample: 2

Alpha: 0,05

Output options:

- ☒ Output Range: \$A\$12
- ☐ New Worksheet Ply:
- ☐ New Workbook

Tỉnh	2003	2004	
Sơn La	4440	4247,7	
Hà Tây	4850	4294,3	
Đắc Lắc	4789	4321,8	
Đồng Nai	4400	4284,3	
	4500	4389,6	
	4500	4314,3	
	4450	4400	

Anova: Two-Factor With Replication

Input Range: \$A\$2:\$C\$10

Rows per sample: 2

Alpha: 0,05

Output options:

- ☒ Output Range: \$A\$12
- ☐ New Worksheet Ply:
- ☐ New Workbook

Source of Variation	SS	df	MS	F	P-value	F crit
Sample	83496,688	3	27832,2	8,85617	0,00637	4,06618
Columns	220195,56	1	220196	70,0658	3,1E-05	5,31766
Interaction	106529,29	3	35509,8	11,2991	0,00301	4,06618
Within	25141,56	8	3142,7			
Total	435363,1	15				

b) Lệnh TOOLS/ DATA ANALYSIS/ T-TEST:

Lệnh này dùng để kiểm định giả thuyết về sự khác nhau giữa 2 số bình quân theo phân phối T, trong EXCEL có 3 kiểu so sánh:

* Lấy mẫu theo cặp Paired two sample for means:

So sánh 2 số bình quân của 2 mẫu khi rút mẫu theo cặp (kiểu cặp đôi) hay người ta còn gọi là so sánh 2 tập hợp số liệu về một chỉ tiêu nhưng ở hai điều kiện khác nhau có số đơn vị mẫu bằng nhau.

Thí dụ: Công ty VINAMILK áp dụng công nghệ mới trong chế biến sữa chua. Hãy kiểm định xem năng suất lao động của công nhân sau khi sử dụng công nghệ mới với công nghệ cũ có khác nhau không với mức ý nghĩa là 5%?

Giải: Lấy mẫu 10 công nhân trong Công ty, thu thập số liệu về năng suất lao động của 10 công nhân này trước và sau khi áp dụng công nghệ mới. Kết quả điều tra thể hiện bảng sau:

Năng suất lao động (NSLĐ) của 10 công nhân điều tra

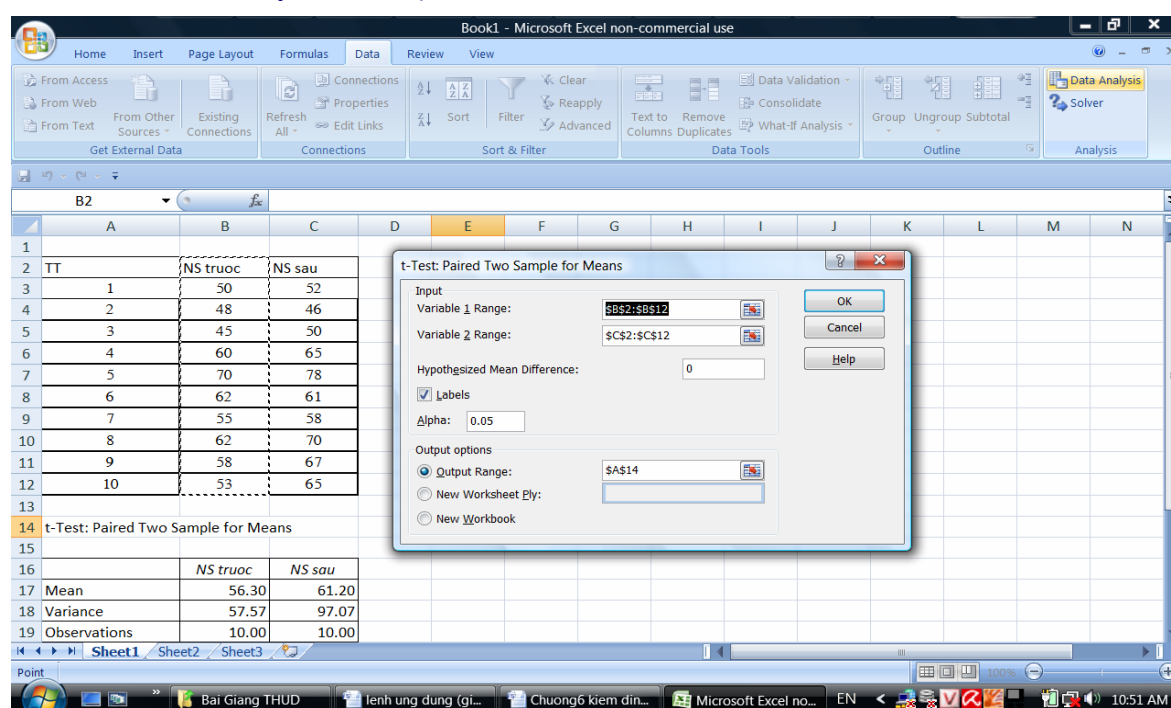
Thứ tự công nhân quan sát	NSLĐ (kg/ngày)	
	Trước khi X	Sau khi Y
1	50	52
2	48	46
3	45	50
4	60	65
5	70	78
6	62	61
7	55	58
8	62	70
9	58	67
10	53	65

Thao tác trên EXCEL như sau:

- Vào Tools/ Data Analysis/ T-Test: Paired two sample for means,
- Input Variable 1 Range: Miền vào của mẫu quan sát thứ 1
- Input Variable 2 Range: Miền vào của mẫu quan sát thứ 2
- Nếu để tên mẫu thì chỉ định Labels
- Nếu muốn kiểm định giả thiết $H_0: m_x = m_y$ thì giả thiết về hiệu số của 2 trung bình ghi số 0.
- Nếu muốn kiểm định giả thiết $H_1: m_x \neq m_y$ (hay muốn kiểm định giả thiết $H_0: m_x = m_y + d$) thì giả thiết về hiệu số của 2 trung bình ghi số d.
- Sau đó chọn miền ra.

OK

Thao tác này minh họa trên màn hình sau:



t-Test: Paired Two Sample for Means

		NS trước	NS sau
Mean	Trung bình của mẫu	56.30	61.20
Variance	Phương sai của mẫu: s_x^2, s_y^2	57.57	97.07
Observations	Số quan sát n_x, n_y	10.00	10.00
Pearson Correlation	Hệ số tương quan r của 2 mẫu (coi như 2 biến)	0.90	
Hypothesized Mean Difference	Giả thuyết H_0	0.00	
df	Bậc tự do (n-1)	9.00	
t Stat	Giá trị t thực nghiệm	-3.46	
P(T<=t) one-tail	Xác suất P(T<t) một phía	0.00	
t Critical one-tail	Giá trị t thực nghiệm một phía	1.83	
P(T<=t) two-tail	P(T<t) hai phía	0.01	
t Critical two-tail	Giá trị t thực nghiệm hai phía	2.26	

Chú ý:

$P(T < t)$ một phía và 2 phía tìm qua hàm $Tdist(z, f, 1)$ và $Tdist(z, f, 2)$

Giá trị T tới hạn một phía tìm qua hàm $Tinv(\alpha, f)$

Giá trị T tới hạn hai phía tìm qua hàm $Tinv(\alpha/2, f)$.

* **Two-sample assuming equal variance:** Lấy mẫu độc lập có số đơn vị mẫu không bằng nhau và giả thiết phương sai bằng nhau.

So sánh 2 số bình quân từ 2 mẫu khi rút mẫu độc lập với giả thiết 2 phương sai bằng nhau có 2 trường hợp:

+ **Trường hợp mẫu lớn:** Khi việc rút mẫu quan sát từ tổng thể được tiến hành một cách độc lập thì chúng ta có 2 mẫu độc lập. Nếu dung lượng 2 mẫu đều lớn (thường quy ước $n_1, n_2 > 30$) ta có thể làm theo Z-test và thay cho 2 phương sai đã biết của tổng thể σ_1^2, σ_2^2 ta dùng 2 phương sai mẫu s_1^2, s_2^2 :

So sánh 2 mẫu độc lập khi biết phương sai thể σ_1^2, σ_2^2 :

Khi rút mẫu độc lập từ 2 tổng thể phân phối chuẩn mà do một tình huống nào đó đã biết được phương sai (thường xảy ra khi điều tra lại một tổng thể sau một thời gian chưa lâu, nên phương sai chưa thay đổi, do đó lấy phương sai của lần điều tra trước làm σ_1^2, σ_2^2 , hoặc khi xem xét các kết quả phân tích của 2 phòng thí nghiệm mà độ chính xác đã ổn định, từ đó có các ước lượng về phương sai).

Thí dụ: Một trại chăn nuôi gà tiến hành thí nghiệm sử dụng 2 loại thức ăn A và B trên cùng một giống. Sau một thời gian thử nghiệm cho ăn, người ta điều tra 35 con nuôi bằng thức ăn A, 35 con nuôi bằng thức ăn B thu được các số liệu về khối lượng trung bình 1 con (kg/con):

TT	Thức ăn A	Thức ăn B	TT	Thức ăn A	Thức ăn B
1	2.3	1.3	19	2.3	1.4
2	2.4	1.5	20	2.2	1.5
3	1.8	1.2	21	2.1	1.8
4	2	1	22	2.4	1.3
5	2.1	0.9	23	2.8	1.4
6	2.5	1.4	24	2.1	1.5
7	2.6	1.5	25	1.9	1
8	2.3	1.6	26	2.7	0.9
9	2.2	1.3	27	2	0.6
10	1.9	1.2	28	2.3	1.6
11	1.5	0.6	29	2.5	1.8
12	2.4	0.8	30	2.3	1.4
13	2.6	0.9	31	2.1	1.5
14	2.3	1.2	32	1.8	1.4
15	2.8	1.3	33	1.8	1.9
16	1.5	1.4	34	1.9	1.3
17	1.7	1.1	35	2.5	1.6
18	1.8	1.4			

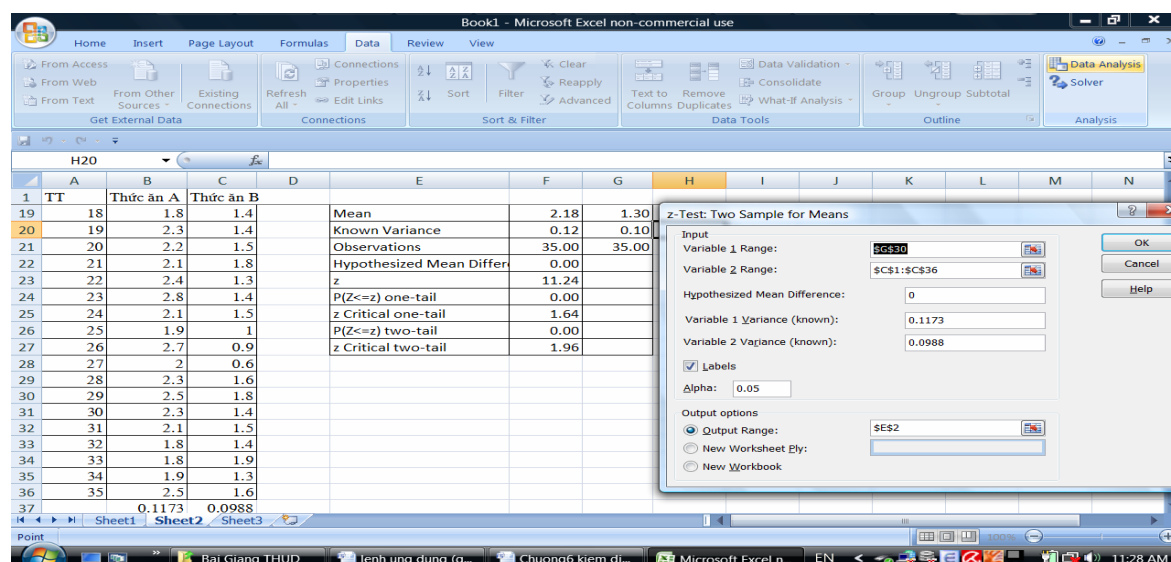
Yêu cầu: Anh chị hãy cho biết khối lượng trung bình 1 con sử dụng ở 2 loại thức ăn sau thời gian nuôi có khác nhau không với mức ý nghĩa là 5%? Biết rằng đợt thí nghiệm tháng trước cho biết phương sai về khối lượng trung bình 1con sử dụng thức ăn A là 0,1173; thức ăn B là 0,0988.

Chúng ta giải bài toán này trên E XCEL với lệnh Z-test for means với các thao tác sau:

- Vào Tools/ Data Analysis/ Z-test for means, sau đó lần lượt trả lời:
- Input Variable 1 Range: Miền vào của mẫu quan sát thứ 1
- Input Variable 2 Range: Miền vào của mẫu quan sát thứ 2

- Nếu để tên mẫu thì chọn Labels
- Nếu muốn kiểm định giả thiết $H_0: m_x = m_y$ thì giả thiết về hiệu số của 2 trung bình ghi số 0.
- Nếu muốn kiểm định giả thiết $H_1: m_x \neq m_y$ (hay muốn kiểm định giả thiết $H_0: m_x = m_y + d$) thì giả thiết về hiệu số của 2 trung bình ghi số d).
- Nhập phương sai σ_1^2 của mẫu quan sát thứ 1
- Nhập phương sai σ_2^2 của mẫu quan sát thứ 2
- Sau đó chọn miền ra.
- OK

Xem minh họa trên hình sau:



Phần kết quả ta được 2 cột cho 2 mẫu quan sát và lần lượt có các hàng:

		Thức ăn A	Thức ăn B
Mean	Số trung bình của mẫu	2.18	1.30
Known Variance	Phương sai cho trước	0.12	0.10
Observations	Số quan sát	35.00	35.00
Hypothesized Mean Difference	Giả thuyết H_0	0.00	
Z	Z thực nghiệm	11.24	
P(Z<=z) one-tail	Xác suất P(Z<z) với giả thiết 1 phía ($H_0: m_x = m_y$)	0.00	
z Critical one-tail	Z tra bảng một phía	1.64	
P(Z<=z) two-tail	Xác suất P(T<t) 2 phía ($H_0: m_x = m_y$)	0.00	
z Critical two-tail	Z tra bảng hai phía	1.96	

- Ta có thể tính lấy một số kết quả:
 - + Tìm Z lý thuyết qua hàm NORMSINV với $\alpha = 0,025$ hai phía: $\alpha = 0,005$ i phía trong EXCEL
 - + P(Z<z) một phía tính bằng hàm 1 - Normsdist(z)
 - + P(Z<Z) hai phía tính bằng hàm (1-Normsdist(z))/ 2
- Tiêu chuẩn kiểm định:
 - + Nếu $|Z| \leq Z_{\alpha/2}$ ta chấp nhận giả thuyết H_0 , coi 2 số trung bình của 2 mẫu bằng nhau
 - + Nếu $|Z| > Z_{\alpha/2}$ ta bác bỏ giả thuyết H_0 , coi 2 số trung bình của 2 mẫu khác nhau
- * **Trường hợp mẫu bé:** Nếu mẫu bé (n_1 hoặc n_2 hoặc cả hai đều < 30. Ta sử dụng lệnh **t-Test: Two-Sample Assuming Equal Variances**

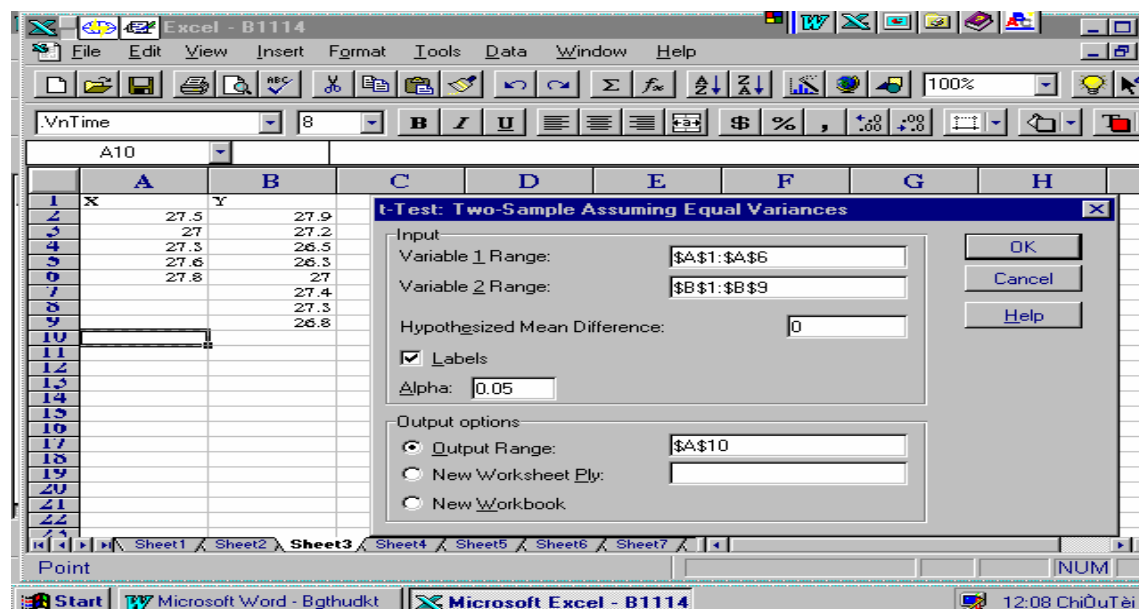
Các thao tác trên EXCEL cũng tương tự như đối với Z-test for means ta nhập riêng mẫu 1 và mẫu 2, nếu có tên thì ghi Labels, giả thiết về hiệu số của 2 trung bình thì ghi 0 nếu muốn kiểm định giả thiết $H_0: \mu_1 = \mu_2$ còn muốn kiểm định giả thiết $H_0: \mu_1 = \mu_2 = d$ thì ghi d. Sau đó chọn miền ra.

Phần kết quả ta được 2 cột cho 2 bảng mẫu quan sát và lần lượt có các hàng:

- Trung bình của mẫu: a_1, a_2
- Phương sai của mẫu: s_1^2, s_2^2
- Số quan sát n_1, n_2 của 2 mẫu.
- Phương sai chung (Pooled variance) tính theo công thức:

$$s^2 = ((n_1 - 1)s_1^2 + (n_2 - 1)s_2^2) / (n_1 + n_2 - 2)$$
- Bậc tự do f để kiểm định giả thiết H_0 , $f = (n_1 + n_2 - 2)$.
- Giá trị t thực nghiệm tính theo công thức: $t = (a_1 - a_2) / \text{SQRT}(s^2 * (1/n_1 + 1/n_2))$
- Xác suất $P(T < t)$ một phía và $P(T < t)$ hai phía.
- Các giá trị tới hạn của hàm T một phía và 2 phía:
 $P(T < t)$ một phía và 2 phía tìm qua hàm Tdist(z, f, 1) và Tdist(z, f, 2).
- Giá trị T tới hạn một phía tìm qua hàm Tinv(α , f)
- Giá trị T tới hạn 2 phía tìm qua hàm Tinv($\alpha/2$, f)

Chúng ta có ví dụ minh họa sau:



Chọn OK ta có kết quả sau:

t-Test: Two-Sample Assuming Equal Variances			
	X	Y	
Mean	27.44	27.05	Trung bình
Variance	0.093	0.265714286	Phương sai
Observations	5	8	Số quan sát
Pooled Variance	0.202909091		Phương sai chung
Hypothesized Mean Difference	0		Giả thiết về hiệu số 2 trung bình
df	11		Bậc tự do
t Stat	1.518700633		t thực nghiệm
P(T<=t) one-tail	0.078521515		P(T<t) một phía
t Critical one-tail	1.795883691		Giá trị t lý thuyết một phía
P(T<=t) two-tail	0.15704303		P(T<t) hai phía
t Critical two-tail	2.200986273		Giá trị t lý thuyết 2 phía

Kết luận: Trung bình 2 mẫu không khác nhau.

*** Two-sample assuming unequal variances:** Hai tập hợp có số đơn vị mẫu không bằng nhau và giả thiết phương sai không bằng nhau (So sánh số bình quân 2 mẫu khi rút mẫu độc lập với giả thiết 2 phương sai khác nhau):

Trường hợp mẫu bé và phương sai khác nhau thì có nhiều cách làm gần đúng, trong EXCEL đã làm theo cách hiện nay hay dùng.

Phần Input Range cũng tương tự như đối với Z-test for means, ta nhập riêng mẫu 1 và mẫu 2, nếu có để tên mẫu thì ghi Labels, giả thiết về hiệu số của 2 trung bình thì ghi 0 nếu muốn kiểm định giả thiết $H_0: \mu_1 = \mu_2$ còn nếu muốn kiểm định giả thiết $H_0: \mu_1 = \mu_2 + d$ thì ghi d. Sau đó chọn miền ra.

Phần kết quả thu được 2 cột cho 2 mẫu quan sát và lần lượt có các hàng:

- Trung bình của mẫu a_1, a_2
- Phương sai của mẫu s_1^2, s_2^2
- Số quan sát n_1, n_2
- Giá trị t thực nghiệm tính theo công thức: $t = (a_1 - a_2) / \text{SQRT}(s_1^2/n_1 + s_2^2/n_2)$
- Nội dung của phép tính gần đúng nằm ở chỗ tính lại bậc tự do f của t theo công thức xấp xỉ sau:

Tính giá trị $u = (s_1^2/n_1) / (s_1^2/n_1 + s_2^2/n_2)$

$$f = 1 / (u^2 / (n_1 - 1) + (1 - u)^2 / (n_2 - 1))$$

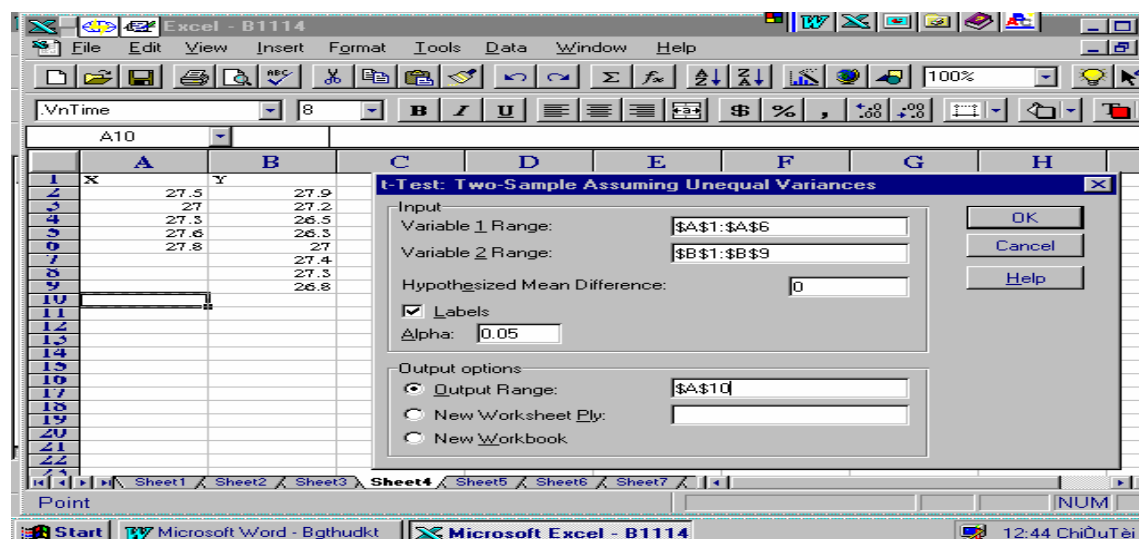
- Các giá trị tới hạn của hàm T một phía và 2 phía:
 - + Giá trị T tới hạn một phía tìm qua hàm $\text{Tinv}(\alpha, f)$
 - + Giá trị T tới hạn 2 phía tìm qua hàm $\text{Tinv}(\alpha/2, f)$
- Xác suất $P(T < t)$ một phía và $P(T < t)$ hai phía:
 - + $P(T < t)$ một phía tìm qua hàm $\text{Tdist}(z, f, 1)$
 - + $P(T < t)$ hai phía tìm qua hàm $\text{Tdist}(z, f, 2)$

Chúng ta có ví dụ minh họa sau (hình trang sau)

Chọn OK ta được kết quả như sau:

t-Test: Two-Sample Assuming Unequal Variances			
	<i>X</i>	<i>Y</i>	
Mean	27.44	27.05	Trung bình
Variance	0.093	0.2657	Phương sai
Observations	5	8	Số quan sát
Hypothesized Mean Difference	0	Giả thiết về hiệu số 2 trung bình	
df	11	bậc tự do	
t Stat	1.7133	t thực nghiệm	
P(T<=t) one-tail	0.0573	P(T<t) một phía	
t Critical one-tail	1.7958	Giá trị t lý thuyết một phía	
P(T<=t) two-tail	0.1146	P(T<t) hai phía	
t Critical two-tail	2.2009	Giá trị t lý thuyết hai phía	

Kết luận: Trung bình 2 mẫu không khác nhau



c) Lệnh TOOLS/ DATA ANALYSIS/ F-TEST: TWO SAMPLE FOR VARIANCES

Lệnh này dùng kiểm định giả thuyết về sự bằng nhau giữa 2 phương sai của 2 tổng thể từ 2 tập hợp số liệu. Giả thiết có 2 mẫu độc lập được chọn ngẫu nhiên có độ lớn n_1 và n_2 . Tổng thể thứ nhất có phân phối chuẩn $N(\mu_1, \delta_1^2)$ tổng thể thứ 2 cũng có phân phối chuẩn với $N(\mu_2, \delta_2^2)$. Các giá trị thống kê trên đều chưa biết mà ta chỉ ước lượng không chệch của phương sai δ_1^2, δ_2^2 là phương sai mẫu s_1^2, s_2^2 .

$$s_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_{i1} - \bar{x}_1)^2 \quad s_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (x_{i2} - \bar{x}_2)^2$$

Trên cơ sở mẫu chọn ta cần đi kiểm định giả thiết $H_0: \delta_1^2 = \delta_2^2$ (2 phương sai coi là đồng nhất). Đối thiết $H_1: \delta_1^2 \neq \delta_2^2$ (2 phương sai không đồng nhất).

Tiêu chuẩn kiểm định phù hợp ở đây được áp dụng là:

F = Phương sai lớn hơn/ Phương sai bé hơn hay $F = s_2^2 / s_1^2$ nếu $s_2^2 > s_1^2$

F gọi là đại lượng ngẫu nhiên Fisher-Snedecor tuân theo luật phân phối Fisher (F) với bậc tự do k_1, k_2 . Bậc tự do $k_1 = n_1 - 1$ và bậc tự do $k_2 = n_2 - 1$

Giá trị F được tính theo biểu thức trên gọi là giá trị thực nghiệm F_t , giá trị tiêu chuẩn (tới hạn) $F_\alpha(k_1, k_2)$ tra ở bảng Fisher-Snedecor.

Nếu $F_t < F_\alpha(k_1, k_2)$ chấp nhận giả thiết $H_0: \delta_1^2 = \delta_2^2$ (2 phương sai là đồng nhất).

Nếu $F_t > F_\alpha(k_1, k_2)$ bác bỏ giả thiết H_0 , chấp nhận đối thiết $H_1: \delta_1^2 \neq \delta_2^2$ (nghĩa là 2 phương sai không đồng nhất).

Thí dụ: Công ty chè Phú Đa sử dụng 2 máy đóng gói chè đen xuất khẩu. Để kiểm tra mức độ chính xác của 2 máy này, người ta chọn mỗi máy 20 túi sản phẩm. Tính toán phương sai về khối lượng trung bình 1 túi cho thấy ở máy 1 là 17 gam/túi, máy 2 là 26 gam/túi. Với mức ý nghĩa là 5% hãy cho biết độ chính xác của 2 máy có như nhau không? ta hãy kiểm định sự bằng nhau của 2 phương sai với 2 tập số liệu sau:

Ta vào Tools/ Data Analysis/ F-test two sample for variance, sau đó trả lời:

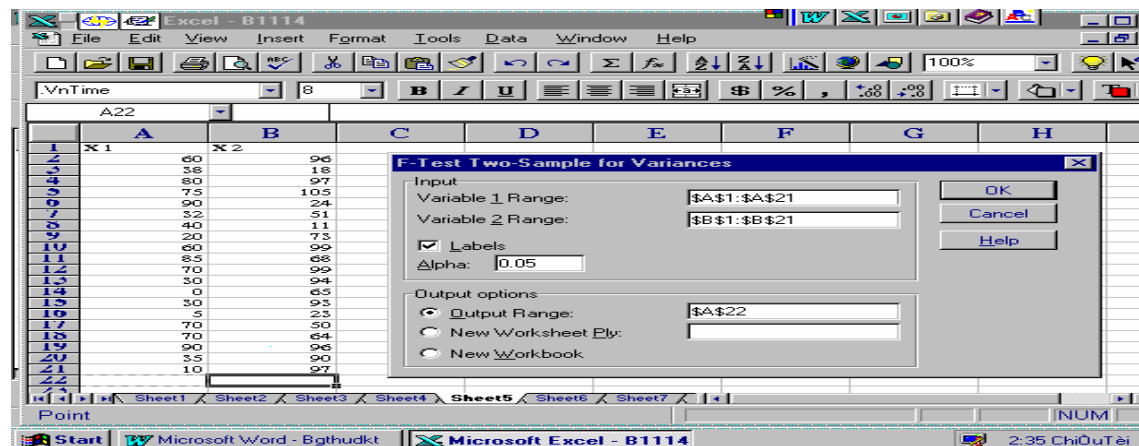
- Input Variable 1 Range: Miền vào của mẫu quan sát thứ 1
- Input Variable 2 Range: Miền vào của mẫu quan sát thứ 2
- Nếu để tên mẫu thì chọn Labels
- Nếu chọn mức ý nghĩa thống kê α thì khai báo, máy mặc nhận là 0,05
- Sau đó chọn miền ra.

Phần kết quả ta được 2 cột cho 2 mẫu quan sát và lần lượt có các hàng:

- Trung bình của mẫu: $\bar{x}_x, \bar{a}_y$ - Phương sai đã cho: σ_1^2, σ_2^2
- Số quan sát n_x, n_y (bằng nhau hoặc không bằng nhau).
- Bậc tự do k_1, k_2 - Giá trị F thực nghiệm.

- Xác suất $P(F < f)$ với giả thiết 1 phía ($H_0: \delta_1^2 = \delta_2^2$ với đối thiết $H_1: \delta_1^2 \neq \delta_2^2$)
- Các giá trị tới hạn của hàm F một phía.

Chúng ta có ví dụ minh họa sau:



Chọn OK để có kết quả:

F-Test Two-Sample for Variances			
	<i>X 1</i>	<i>X 2</i>	
Mean	49.5	70.65	Trung bình
Variance	840.1578947	976.7657895	Phương sai
Observations	20	20	Số quan sát
df	19	19	Bậc tự do
F	0.86014263		F thực nghiệm
P(F<=f) one-tail	0.626996283		P(F<F _{tn}) 1 phía
F Critical one-tail	0.461200855		F tới hạn 1 phía

Chú ý:

- Vì ta lấy **phương sai lớn chia cho phương sai nhỏ** được F_{tn} nên ở đây ta tính được $F_{tn} = 976,766 / 840,1579 = 1,162598$ (đối chiếu với kết quả máy chạy là $840,1579 / 976,766 = 0,860143$).

- Vì chúng ta phải tính lại F_{tn} , do đó tất yếu phải tính lại $P(F > F_{tn})$ và F tới hạn:

+ $P(F > F_{tn})$ tính bằng hàm **Fdist(F_{tn} @ f tử số @ f mẫu số) = 0.369740299**

+ F tới hạn tính bằng hàm **Finv(α @ f tử số @ f mẫu số) = 2.124153298**

Căn cứ vào kết quả tính lại (số đậm) ta kiểm định giả thiết và đi đến kết luận rằng hai phương sai không khác nhau ở mức $\alpha = 0,05$.

d) Lệnh TOOLS/ DATA ANALYSIS/ REGRESSION: Dùng để giải bài toán hồi quy tuyến tính đơn và hồi quy bội.

*** Giới thiệu chung:**

Hiện nay có rất nhiều chương trình có thể giải bài toán dạng tương quan hồi quy (tuyến tính hay phi tuyến) nhưng trong kinh tế thường áp dụng các chương trình LIMDEP, SHAZAM, EXCEL (LOTUS và QUATTRO cũng có thể giải được loại toán này nhưng kết quả bị hạn chế), SPSS, POM v.v... Tuy nhiên, chương trình EXCEL vẫn thông dụng hơn cả bởi vì các thao tác tương đối đơn giản rất phù hợp với điều kiện người học. Khi SETUP chương trình ban đầu cần khai báo phần ANALYSIS TOOKPAK thì khi đó mới có thể giải được những bài toán này.

*** Các thao tác chính:**

Vào dữ liệu bình thường, bố trí theo cột - mỗi cột là 1 chỉ tiêu, giữa cột chỉ tiêu phân ánh kết quả và đầu vào có thể không liền kề.

Lệnh chính: Chọn TOOLS/ DATA ANALYSIS với EXCEL 5.0 và chọn OPTIONS/ ANALYSIS TOOLS với EXCEL 4.0 hộp hội thoại ANALYSIS sẽ xuất hiện (có nhiều lệnh

giải các bài toán thống kê: Tương quan (CORRELATION), số bình quân di động (MOVING AVERAGE), T-TEST, ANOVA, ... Chọn (bấm) theo danh sách (theo thanh cuộn hoặc theo chiều mũi tên lên, xuống), chọn mục REGRESSION (hồi quy), sau đó OK hay bấm đúp chuột tại mục REGRESSION.

Hộp hội thoại REGRESSION xuất hiện. Trong đó cần phải khai báo địa chỉ của:

- **Input Y range:** Miền chứa kết quả (cột kết quả - chỉ được 1 cột, có thể chứa cả tiêu đề). Đây là biến phụ thuộc Y.

- **Input X range:** Miền dữ liệu phản ánh cá biến độc lập mà ta quy định ban đầu (có thể là 1 hoặc nhiều cột). Đây chính là các biến độc lập X_i .

Nếu miền Y và X_i ta khai báo địa chỉ có cả tiêu đề thì mục Labels phải đánh dấu.

Khai báo địa chỉ để EXCEL đưa kết quả ra - nên đưa ngay ở bảng tính hiện thời.

- Bấm chuột tại **Output range** sau đó khai báo địa chỉ - là miền trống đủ rộng để chứa kết quả (9 cột, 18 hàng tối thiểu, nếu thêm biến thì thêm hàng).

- Ngoài ra còn các mục khác - ta có thể khai báo hoặc không (khi đó máy sẽ chạy theo mặc nhận - tối ưu nhất). Ví dụ:

+ **Residuals:** Là phần dư khi tính toán (hiệu số giữa giá trị Y thực nghiệm và giá trị tính toán theo hồi quy (Predicted values)). Nếu mô hình tốt thì phần dư này phải phân bố ngẫu nhiên theo luật chuẩn, điều này thể hiện rõ nếu ta lấy phần dư đã chuẩn hoá (Standartized residuals).

+ **Confidence level:** Là mức độ tin cậy (mặc nhiên là 95%).

+ Nếu ta chọn Line fit plot thì EXCEL sẽ tính số đường hồi quy bằng số biến độc lập theo mô hình hồi quy tuyến tính đơn.

Nếu chọn Residual Plot thì máy sẽ vẽ phần dư của các đường hồi quy đơn nói trên. Ta có thể thấy các phần dư tản mạn ngẫu nhiên 2 bên đường trục ngang và có biên độ ngẫu nhiên (luật chuẩn) thì tốt, nếu phần dư cụm lại đoạn (toàn ở trên, đoạn khác toàn ở dưới thì phải xem lại số liệu và nếu cần thì có thể biến đổi theo một số biến đổi thường dùng.

+ Nếu chọn Normal Probability Plot thì máy sẽ theo dõi các biến phụ thuộc Y và căn cứ vào luật chuẩn để vẽ các điểm ghi lại Y_i từ nhỏ tới lớn, nếu các điểm nằm trên đường thẳng thì chấp nhận giả thiết biến Y tuân theo luật chuẩn.

- Sau khi khai báo xong, bấm OK máy sẽ bắt đầu giải và cuối cùng cho:

+ Bảng thống kê cơ bản (Summary statistics), trong đó có tính hệ số tương quan R, hệ số tương quan bình phương R^2 (còn gọi là hệ số tương quan xác định), hệ số tương quan bình phương sau khi hiệu chỉnh nếu mẫu nhỏ (số quan sát n ít), sai số chuẩn Se (căn bậc hai của giá trị Se^2 ở trong bảng phân tích phương sai ở dòng biến động do Error, cuối cùng là số quan sát n.

+ Bảng phân tích phương sai (ANOVA): Bảng này tách toàn bộ biến động (TSS) ra thành 2 thành phần: Biến động do hồi quy (RSS) và biến động do sai số (ESS), mỗi biến động có số bậc tự do tương ứng.

Căn cứ vào giá trị F thực nghiệm để kết luận mô hình hồi quy tìm được có đáng tin cậy hay không. Muốn kết luận phải so sánh F thực nghiệm với F lý thuyết tìm được với mức sai alpha (thí dụ 0,05), bậc tự do của tử số là bậc tự do của phần biến động do hồi quy (bảng số biến độc lập k), bậc tự do của mẫu số là bậc tự do của sai số (bảng $n - 1 - k$). cũng có thể căn cứ vào giá trị xác suất ở cột bên cạnh để kết luận, nếu giá trị p đó bé hơn alpha (0,05) thì hồi quy đáng tin cậy, nếu lớn hơn alpha thì hồi quy không đáng tin cậy.

+ Bảng tiếp theo cho ta các hệ số của các biến trong phương trình hồi quy lần lượt ta có hằng số a_0 , hệ số a_1 của biến 1, hệ số a_2 của biến 2, v.v....

Cột bên cạnh tính sai số của các hệ số. Nếu lấy giá trị của các hệ số chia cho sai số tương ứng thì ta được giá trị t thực nghiệm. So sánh giá trị tuyệt đối của t thực

thực nghiệm với t lý thuyết tìm được với sai số alpha và bậc tự do của sai số $(n-1)$ (thông qua hàm TINV) ta có thể đánh giá hệ số nào khác không còn hệ số nào khác không rõ rệt và từ đó có thể đơn giản bớt phương trình hồi quy. Cũng có thể không cần so sánh mà tìm giá trị xác suất ở cột bên cạnh, nếu xác suất đó nhỏ hơn mức alpha (0,05) thì hệ số khác 0 rõ rệt, nếu xác suất lớn hơn alpha thì không khác 0 rõ rệt.

Sau khi phân tích nên chạy lại hồi quy với số biến độc lập ít hơn vì chỉ giữ lại các biến có hệ số khác không rõ rệt.

Hai cột cuối cùng cho ta cận dưới và cận trên của khoảng ước lượng của hệ số. Muốn có cận đó có thể sử dụng công thức $m \pm Se \cdot t$ (với t là giá trị t lý thuyết vừa nói ở trên).

Nếu chọn Option Residuals thì ta được bảng so sánh giá trị Y thực nghiệm và Y tính qua hàm hồi quy (Predicted) gọi là phần dư. Nếu chọn Option Standartized thì được các phần dư chuẩn hoá (trừ đi trung bình và chia độ lệch chuẩn).

Nếu chọn Option Residual plot thì được các hình vẽ các phần dư của các hồi quy tuyến tính đơn của Y theo từng biến $X_1, X_2, \dots, X_k$.

Nếu chọn Option Line fit plot thì được các hình vẽ hồi quy tuyến tính đơn Y theo X_1, Y theo $X_2, \dots, Y$ theo X_k .

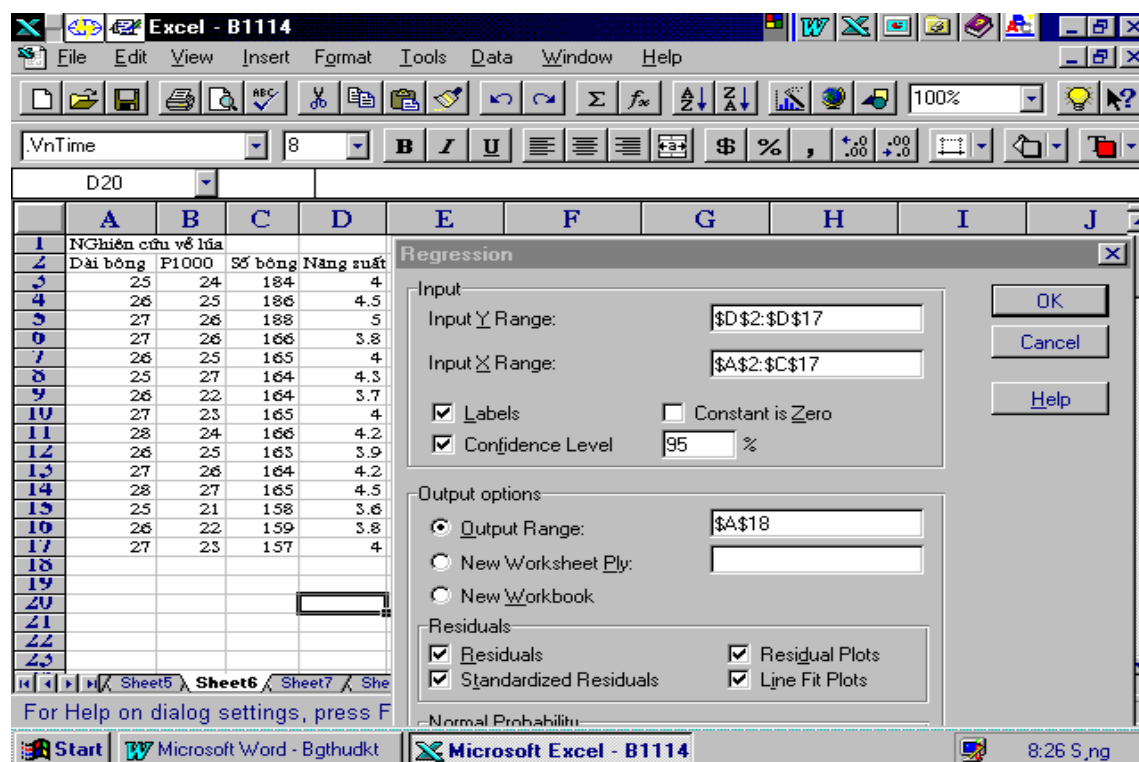
Ta có thể vào chế độ đồ hoạ (Graph) để sửa sang các hình cho đẹp và nếu cần thì thêm phương trình hồi quy tuyến tính và hệ số tương quan R vào hình vẽ.

Trong phần hồi quy của EXCEL không có hồi quy đa thức, nếu muốn làm hồi quy đa thức: $Y = a_0 + a_1X + a_2X^2 + \dots + a_kX^k$ có 2 cách:

+ Làm hồi quy tuyến tính đơn Y theo X sau đó vào chế độ đồ hoạ, xoá các điểm dự báo (Predicted) bằng cách bật sáng dãy điểm dự báo (nháy vào một điểm) rồi gõ phím Del, tiếp theo là bật sáng dãy điểm thực nghiệm rồi vào Menu Insert để Insert trend line, trong hộp hội thoại tiếp theo ta chọn dạng hồi quy là Polynomial bậc k , chọn Option thêm Equation và R^2 sau đó OK.

+ Trong bảng số liệu sau cột X ta thêm ngay các cột $X^2, X^3, \dots, X^k$ sau đó làm hồi quy bội tuyến tính Y theo các biến $X, X^2, X^3, \dots, X^k$. Muốn vậy như vậy trong Input Range của biến độc lập ta phải khai miền bao trùm lên tất cả các biến từ X đến X^k .

Để minh hoạ chúng ta nghiên cứu năng suất lúa Y và 3 biến độc lập (X_i) là chiều dài bông, số bông và trọng lượng nghìn hạt:



Chọn OK chúng ta có kết quả sau:

SUMMARY OUTPUT - Các thống kê cơ bản

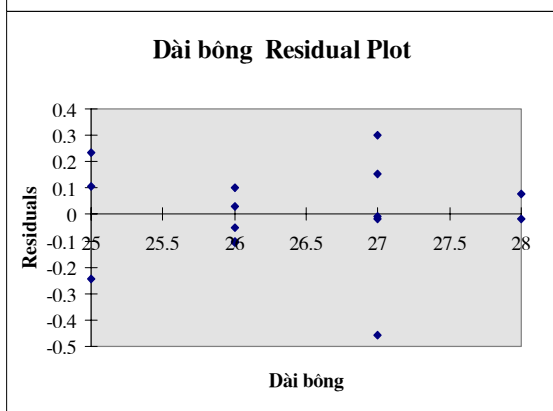
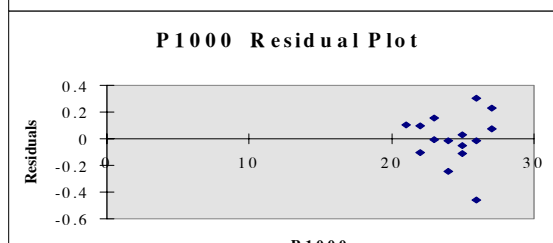
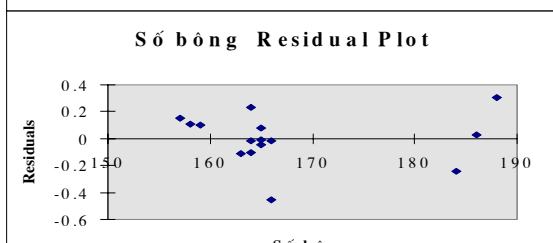
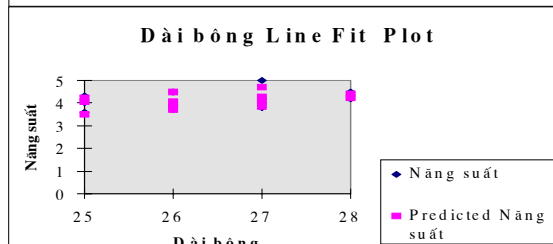
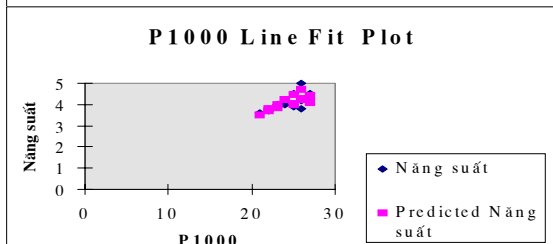
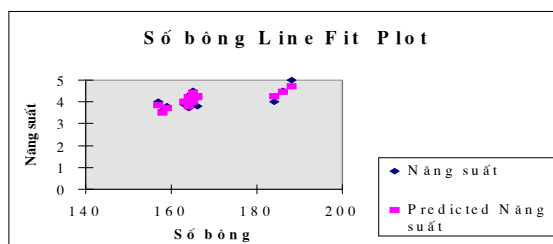
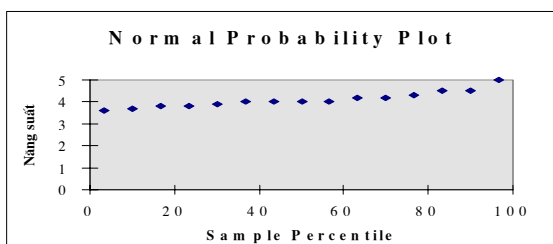
Regression Statistics		Thống kê hồi quy
Multiple R	0.85890092	Hệ số tương quan bội R
R Square	0.73771079	Hệ số tương quan bội bình phương R^2
Adjusted R Square	0.666177369	Hệ số tương quan bội bình phương sau khi hiệu chỉnh R^2
Standard Error	0.210596109	Sai số chuẩn
Observations	15	Số quan sát (Số hàng số liệu)

ANOVA- Bảng phân tích phương sai				F thực nghiệm	Xác suất
	df(bậc tự do)	SS(Sai số b/phg)	MS(TB sai số b/phg)	F	Significance F
Regression (Hồi quy)	3	1.372142069	0.4573807	10.3128129	0.001582207
Residual (Sai số)	11	0.487857931	0.0443507		
Total (Toàn bộ)	14	1.86			

	Hệ số	Sai số chuẩn	Giá trị t	Xác suất	Cánh dưới	Cánh trên
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept (Hằng số)	-4.06364336	1.844184814	-2.2034903	0.04978167	-8.122668822	-0.0046179
Dài bông (Biến 1)	0.111607804	0.060811857	1.8352968	0.09362082	-0.022238258	0.245453865
P1000 (Biến 2)	0.075683509	0.033787699	2.2399723	0.04670353	0.001317247	0.15004977
Số bông (Biến 3)	0.0201105	0.006123395	3.2842076	0.00727919	0.006632992	0.033588007

RESIDUAL OUTPUT (Sai số)				PROBABILITY OUTPUT		
Sai số	Giá trị dự báo	Sai số	Sai số chuẩn hoá	Phân phối của NS với giả thiết phân phối chuẩn		
Observation	Predicted Năng suất	Residuals	Standard Residuals	Percentile	Năng suất	
1	4.243287865	-0.243287865	-1.1552344	3.333333333		3.6
2	4.470800176	0.029199824	0.1386532	10		3.7
3	4.698312488	0.301687512	1.4325408	16.66666667		3.8
4	4.255881497	-0.455881497	-2.1647195	23.33333333		3.8
5	4.048479685	-0.048479685	-0.2302022	30		3.9
6	4.068128399	0.231871601	1.1010251	36.66666667		4
7	3.801318659	-0.101318659	-0.4811041	43.33333333		4
8	4.008720471	-0.008720471	-0.0414085	50		4
9	4.216122283	-0.016122283	-0.0765555	56.66666667		4
10	4.008258686	-0.108258686	-0.5140583	63.33333333		4.2
11	4.215660497	-0.015660497	-0.0743627	70		4.2
12	4.423062309	0.076937691	0.3653329	76.66666667		4.3
13	3.493364349	0.106635651	0.5063515	83.33333333		4.5
14	3.700766161	0.099233839	0.4712045	90		4.5
15	3.847836474	0.152163526	0.7225372	96.66666667		5
9	4.216122283	-0.016122283	-0.0765555	56.66666667		4
10	4.008258686	-0.108258686	-0.5140583	63.33333333		4.2

11	4.215660497	-0.015660497	-0.0743627	70	4.2
12	4.423062309	0.076937691	0.3653329	76.66666667	4.3
13	3.493364349	0.106635651	0.5063515	83.33333333	4.5
14	3.700766161	0.099233839	0.4712045	90	4.5
15	3.847836474	0.152163526	0.7225372	96.66666667	5



Ta được 3 đường hồi quy đơn Y theo X_1 , Y theo X_2 và Y theo X_3 và 3 đường ghi lại phần dư của 3 đường hồi quy trên. Nếu phần dư tản mạn một cách ngẫu nhiên quanh trục ngang thì phần dư được coi như phân phối chuẩn, nếu không như vậy thì phần dư không theo phân phối chuẩn và có sự vi phạm về điều kiện để áp dụng mô hình hồi quy. Khi đó nếu cần có thể biến đổi số liệu bằng các phép biến đổi thường dùng trong thống kê như lấy Log, lấy căn, biến đổi arcsin.

Đường Normal probability plot theo dõi biến năng suất, nếu các điểm nằm trên 1 đường thẳng thì năng suất được coi là phân phối chuẩn.

Tất cả các điều nói trên chỉ đúng khi số liệu tương đối nhiều, còn khi số liệu ít thì chỉ xem các hình vẽ trên và cả các tính toán hồi quy chỉ là các minh họa chứ chưa có giá trị khoa học.

Chú ý:

- Để vẽ đúng các đồ thị trên chúng ta phải đổi cột Năng suất (biến phụ thuộc) đứng ở cột 1, các cột còn lại là vị trí các biến độc lập.
- Mặc dù khi chúng ta khai vị trí của biến phụ thuộc và biến độc lập như minh họa trên thì kết quả trả lời hoàn toàn đúng, song nếu không đặt lại vị trí cột NS đứng ở cột đầu tiên thì đồ thị sẽ sai mà nhiều khi chúng ta không để ý.
- Dù bài toán hồi quy đơn hay đa biến tốt nhất bạn nên đặt biến phụ thuộc (Y) ở cột đầu tiên, các cột còn lại đặt biến độc lập (X_i).

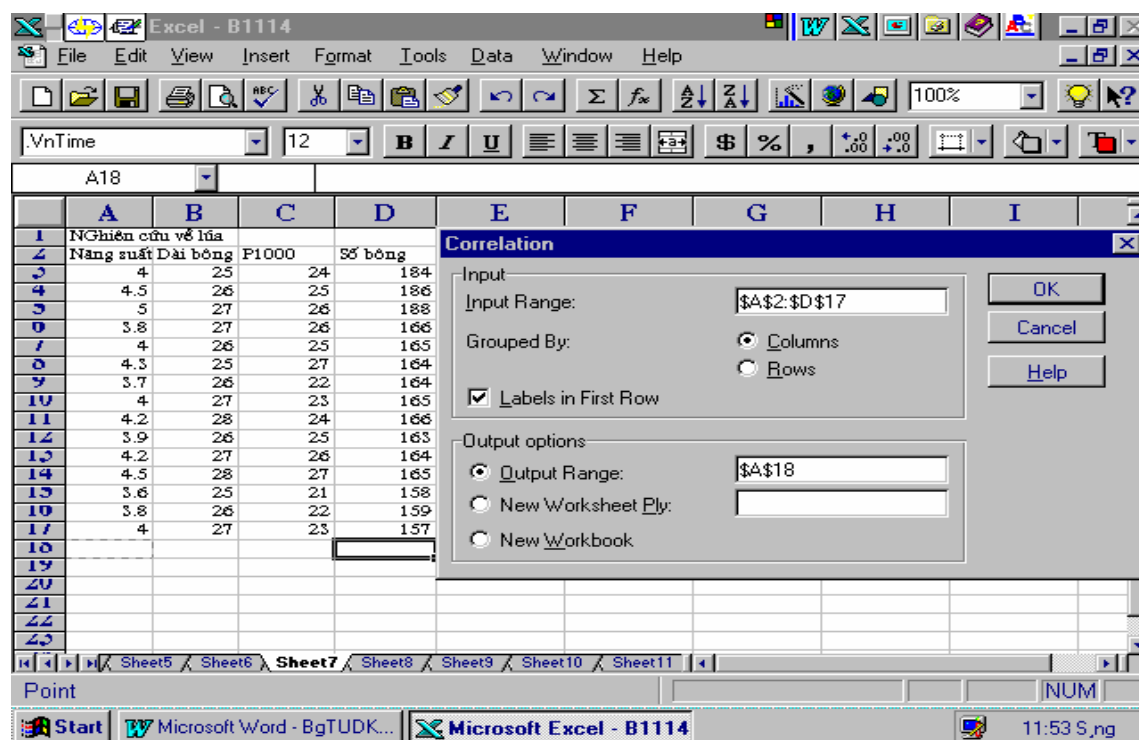
e) Lệnh TOOLS/ DATA ANALYSIS/ CORRELATION

Lệnh này dùng để tính toán ma trận hệ số tương quan riêng (đơn giản) giữa các yếu tố. Phần này tính cho chúng ta bảng các hệ số tương quan đơn giữa các biến sắp xếp thành

một bảng gồm n hàng và m cột, mỗi cột là một biến. Miền vào là toàn bảng kể cả hàng tên biến.

Kết quả ra là bảng các hệ số tương quan đơn giữa các biến.

Chúng ta có thể thấy tác dụng của chuỗi lệnh này qua thí dụ minh họa sau:



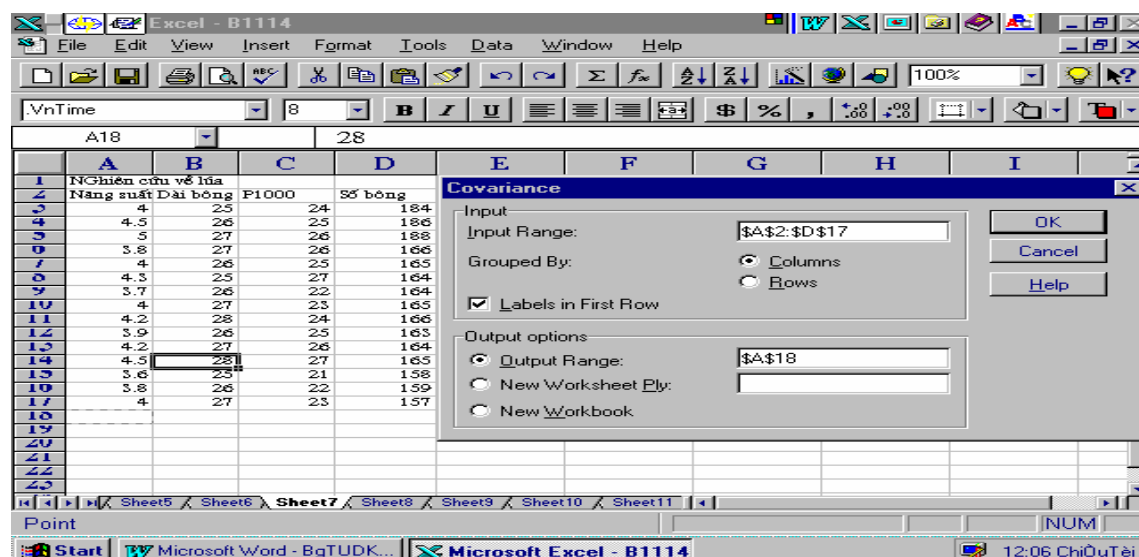
Chọn OK chúng ta có kết quả sau:

	<i>Năng suất</i>	<i>Dài bông</i>	<i>P1000</i>	<i>Số bông</i>
<i>Năng suất</i>	1			
<i>Dài bông</i>	0.37777	1		
<i>P1000</i>	0.66632	0.29262	1	
<i>Số bông</i>	0.66138	-0.06988	0.340772	1

f) Lệnh TOOLS/ DATA ANALYSIS/ COVARIANCE

Lệnh này dùng để tính toán ma trận gồm các phương sai (trên đường chéo) và các hiệp phương sai (các phần tử ngoài đường chéo) của các biến. Phần này vào tương tự như phân hệ số tương quan. Miền vào là toàn bảng kể cả hàng tên biến.

Kết quả ra là bảng các phương sai trên đường chéo và các hiệp phương sai nằm ngoài đường chéo của các biến. Chúng ta có thể thấy tác dụng của chuỗi lệnh này qua thí dụ minh họa sau:



Chọn OK chúng ta có kết quả sau:

	Năng suất	Dài bông	P1000	Số bông
Năng suất	0.13286			
Dài bông	0.13571	0.971429		
P1000	0.45714	0.542857	3.542857	
Số bông	2.4	-0.68571	6.385714	99.11428571

g) Lệnh TOOLS/ DATA ANALYSIS/ DESCRIPTIVE STATISTICS (Thống kê mô tả)

Lệnh này dùng để xác định các chỉ tiêu thống kê đơn giản của các chỉ tiêu: Trung bình, Mode, trung vị, độ lệch chuẩn, phương sai, mô men bậc 3, mô men bậc 4, T-TEST, v.v...

Số liệu để ở trong một bảng có m cột, mỗi cột là một biến, số quan sát có thể không bằng nhau do đó trong phần Input phải chú ý vào miền chữ nhật bao trùm lên toàn bộ số liệu. Vào số liệu sau đó bạn vào Tools/ Data Analysis/ Descriptive statistics, phần đối thoại gồm:

- Input Range: Nhập toàn bộ dữ liệu kể cả dòng có tên biến.
- Labels in First row: x
- Confidence level for means: 95%
- K-th Largest: 1 Nếu muốn chọn số to thứ nhì thì ghi 2
- K-th smallest: 1 Nếu muốn chọn số nhỏ thứ nhì (số nhỏ sau số nhỏ nhất).
- Output Range: Miền ra
- Summary Statistics: x (ở đây nhất thiết phải ghi x thì mới có được các thống kê cơ bản).

Các thống kê đối với mỗi biến được in trên 2 cột, cột đầu tiên ghi tên thống kê, cột sau ghi giá trị của thống kê đó. Lần lượt từ trên xuống dưới chúng ta có:

- + Count: Dung lượng mẫu n (số quan sát).
- + Sum: Tổng các giá trị quan sát theo cột.
- + Average: Trung bình cộng số học giản đơn.
- + Median (Me): Số trung vị (đứng giữa nếu sắp xếp các quan sát từ nhỏ đến lớn).
- + Mode (Mo): Giá trị ứng với tần số cao nhất.
- + Variance: Phương sai mẫu đã hiệu chỉnh

(Phương sai mẫu hiệu chỉnh: $s'^2 = \frac{\sum (x_i - \bar{x})^2}{(n-1)}$; Phương sai mẫu chưa hiệu

chỉnh: $s^2 = \frac{\sum (x_i - \bar{x})^2}{n}$ hay còn gọi là phương sai tổng thể con. Khi $n \rightarrow \infty$ thì s^2 chính là phương sai tổng thể).

+ Standart deviation: Độ lệch chuẩn là căn bậc 2 của phương sai mẫu (s)

+ Standart error: Là sai số của trung bình cộng, bằng độ lệch chuẩn chia cho căn bậc 2 của dug lượng mẫu (s/Qqrt(n)).

+ Kurtosis: Là độ nhọn dùng để đánh giá mật độ phân phối của dãy quan sát có nhọn hơn hay tù hơn đường mật độ chuẩn tắc. Độ nhọn tính theo công thức:

$$\text{Kurtosis} = \frac{1}{(n-1)(n-2)} \sum \left(\frac{x_i - \bar{x}}{s} \right)^3$$

+ Skewness: Là độ bất đối xứng dùng để đánh giá đường mật độ phân phối lệch sang trái (tập trung nhiều về bên trái) hay lệch bên phải (tập trung nhiều về bên phải). Độ bất đối xứng tính theo công thức sau:

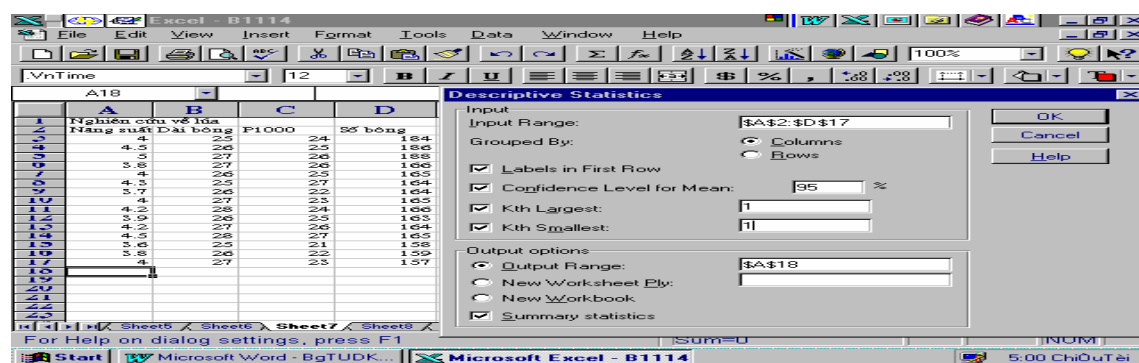
$$\text{Skewness} = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{x_i - \bar{x}}{s} \right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$$

+ Confidence Level: Mức tin cậy P (90%, 95%, 99%) và bằng 1- alpha, với alpha là mức sai (1%, 5%, 10%)

Thống kê tính ở đây là $m = t(\alpha, n-1) * \sigma / \text{Sqrt}(n)$ với $t(\alpha, n-1)$ là giá trị t ở mức sai alpha và n-1 bậc tự do. Nếu so với khoảng tin cậy ở mức $P=1-\alpha$ thì chính là nửa khoảng tin cậy và thường dùng để viết khoảng tin cậy của trung bình cộng dưới dạng: $\bar{x} \pm m$

Trong phần giải thích hàm Confidence Excel đưa ra nội dung Confidence($\alpha=0,05, \sigma$) bằng $1,96 * \sigma / \text{Sqrt}(n)$, nhưng đó chính là nửa khoảng tin cậy khi ước lượng trung bình của tổng thể phân phối chuẩn khi đã biết phương sai σ^2 (Ta có khoảng tin cậy $\bar{x} \pm 1,96 * \sigma / \sqrt{n}$). Vì không biết phương sai σ^2 nên ta dùng $t(\alpha, n-1) * s / \text{Sqrt}(n)$.

Ta có thể minh hoạ qua ví dụ sau:



Chọn OK ta có kết quả sau:

Năng suất		Đài bông		P1000		Số bông	
Mean	4.1	Mean	26.4	Mean	24.4	Mean	167.6
Standard Error	0.09411	Standard Error	0.254484	Standard Error	0.485994317	Standard Error	2.570528943
Median	4	Median	26	Median	25	Median	165
Mode	4	Mode	26	Mode	25	Mode	165
Standard Deviation	0.3645	Standard Deviation	0.985611	Standard Deviation	1.882247896	Standard Deviation	9.955615788
Sample Variance	0.13286	Sample Variance	0.971429	Sample Variance	3.542857143	Sample Variance	99.11428571

Kurtosis	1.30721	Kurtosis	-0.810487	Kurtosis	-0.97494597	Kurtosis	0.569278606
Skewness	1.04158	Skewness	0.061978	Skewness	-0.30255255	Skewness	1.346426593
Range	1.4	Range	3	Range	6	Range	31
Minimum	3.6	Minimum	25	Minimum	21	Minimum	157
Maximum	5	Maximum	28	Maximum	27	Maximum	188
Sum	61.5	Sum	396	Sum	366	Sum	2514
Count	15	Count	15	Count	15	Count	15
Largest(1)	5	Largest(1)	28	Largest(1)	27	Largest(1)	188
Smallest(1)	3.6	Smallest(1)	25	Smallest(1)	21	Smallest(1)	157
Confidence	0.20185	Confidence	0.545814	Confidence	1.042355069	Confidence	5.513241162
Level(95.0%)	Level(95.0%)	Level(95.0%)	Level(95.0%)	Level(95.0%)	Level(95.0%)	Level(95.0%)	Level(95.0%)

h) **Lệnh TOOLS/ DATA ANALYSIS/ HISTOGRAM (Tổ chức đồ)**

Phần này cốt để chia nhóm khi có nhiều số liệu. Ta tìm Max, Min sau đó căn cứ vào biên độ để quyết định chia làm 20-30 nhóm. Trên cơ sở đó tìm bề rộng của nhóm ($h = \text{Biên độ} / \text{Số nhóm}$), sau đó quy tròn h thành 1 số không có số lẻ (số lớn hơn số lẻ của các quan sát).

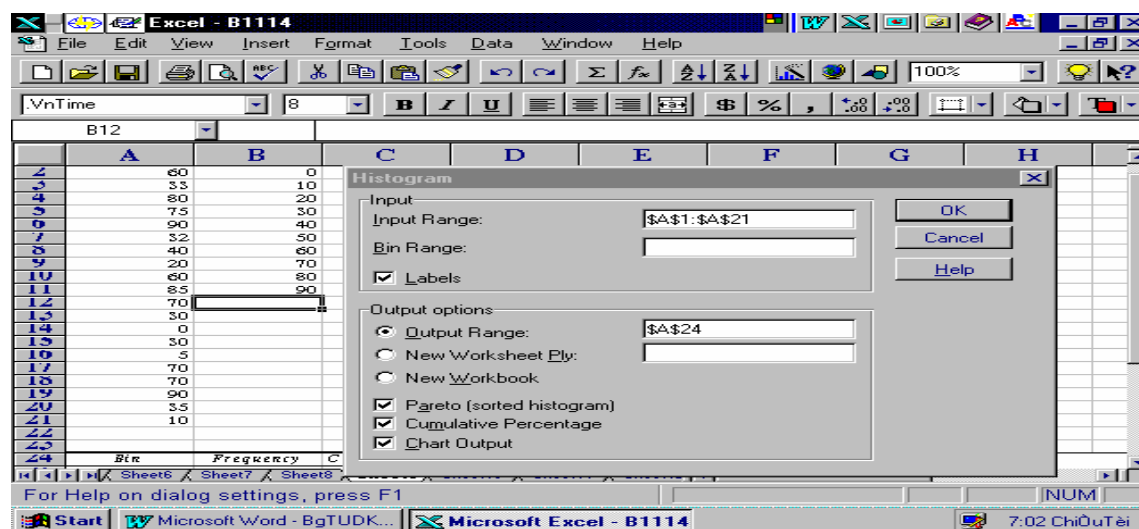
Nếu để máy tự chia thì không cần phải lập thêm cột Bin range, lúc này máy sẽ căn cứ vào Min, Max, Range để chia lấy thành 1 số nhóm, nhưng nếu muốn tự làm thì chọn 1 cột (có thể đặt tên là Bin1, Bin2, ...) sau đó để giá trị Min hoặc 1 giá trị lớn hơn nhưng khá gần giá trị Min ở ô đầu, ô tiếp theo cộng thêm h , ô tiếp theo cộng 2 h , cho đến khi bằng hay vượt Max.

Vào Histogram khai Input range, Bin range (nếu muốn tự làm), Output range.

Option ra có: Pareto (Tính tần số và sắp xếp lại từ nhỏ tới lớn)

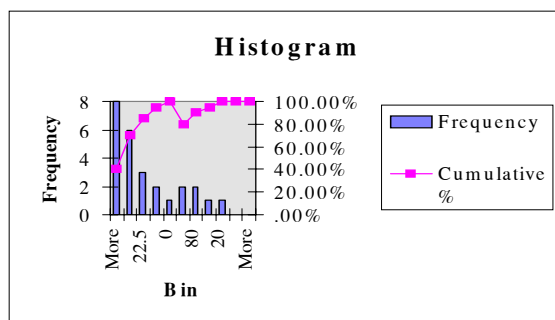
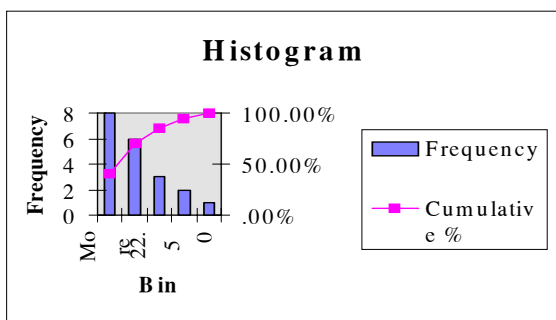
Cumulative (Cộng dồn), Percentage (Tính %).

Ta thu được tần số trong các khoảng và 1 đồ thị vừa vẽ tần số vừa vẽ tần số cộng dồn. Ta xét thí dụ sau: Để máy tự làm Bin range sau đó chọn ra 1 Bin range khác căn cứ vào kết quả ở phần thống kê mô tả.



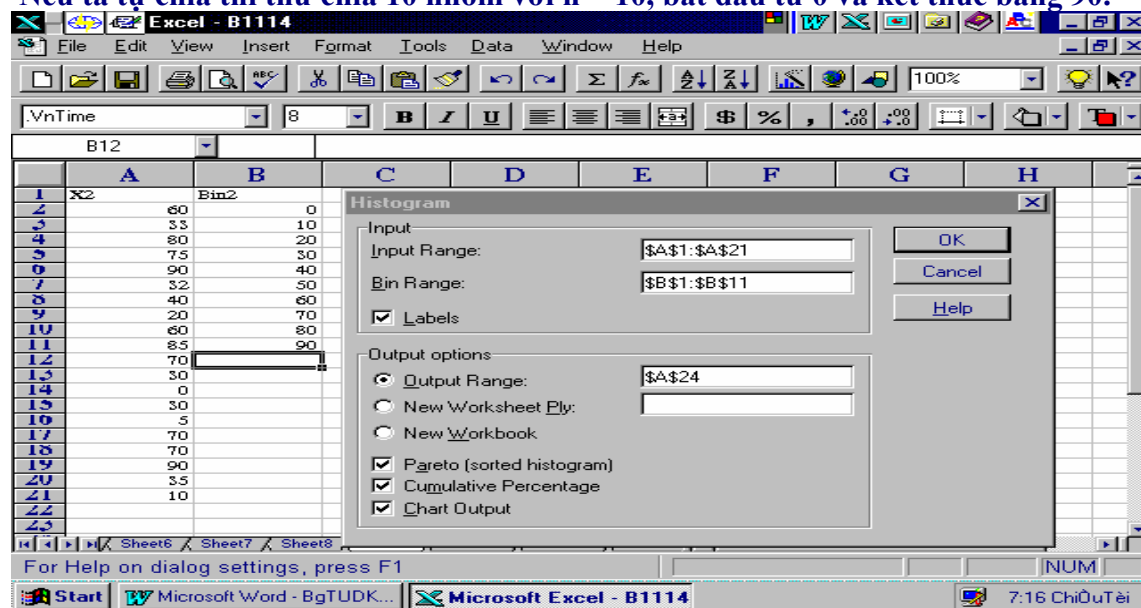
Chọn OK ta có kết quả sau:

Bin	Frequency	Cumulative %	Bin	Frequency	Cumulative %
0	1	5.00%	More	8	40.00%
22.5	3	20.00%	45	6	70.00%
45	6	50.00%	22.5	3	85.00%
67.5	2	60.00%	67.5	2	95.00%
More	8	100.00%	0	1	100.00%
50	0	50.00%	60	2	80.00%



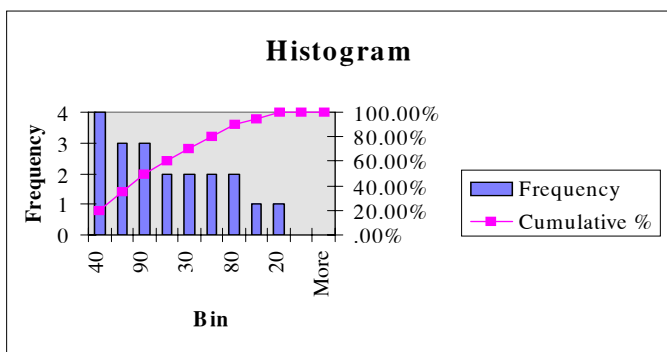
Máy tự chia thành 4 khoảng với bề rộng 22,5.

Nếu ta tự chia thì thử chia 10 nhóm với h = 10, bắt đầu từ 0 và kết thúc bằng 90.



Chọn OK ta có kết quả sau:

Bin	Frequency	Cumulative %	Bin	Frequency	Cumulative %
0	1	5.00%	40	4	20.00%
10	2	15.00%	70	3	35.00%
20	1	20.00%	90	3	50.00%
30	2	30.00%	10	2	60.00%
40	4	50.00%	30	2	70.00%
50	0	50.00%	60	2	80.00%
60	2	60.00%	80	2	90.00%
70	3	75.00%	0	1	95.00%
80	2	85.00%	20	1	100.00%
90	3	100.00%	50	0	100.00%
More	0	100.00%	More	0	100.00%



Đây là thí dụ minh họa còn trong thực tế khi có hàng trăm số liệu thì nên chia từ 20-30 nhóm. Chia nhóm có lợi là cho ta phác họa về dãy phân phối để xem có tuân theo luật chuẩn hay không và sau này khi xử lý thống kê có thể bỏ qua không quan tâm đến dãy số ban đầu nữa mà coi như ta có dãy số x_i (là điểm giữa của các khoảng) với tần số f_i tìm được ở cột tần số. Việc tính trung bình tương đối chính xác còn khi tình phương sai thì nếu thấy cần thiết có thể dùng hiệu chỉnh Se-pốt (Shepard).

2.2. Các lệnh tổng hợp và xử lý số liệu

a) Tính tổng các nhóm số liệu

Trên một CSDL, chúng ta thường có nhu cầu tổng hợp số liệu theo từng nhóm. Ví dụ như theo đơn vị, theo loại vật tư, theo một tài khoản trong kế toán... EXCEL có khả năng phân tích các nhóm dữ liệu, đồng thời chèn vào cuối mỗi nhóm những dòng thống kê các giá trị cần thiết. Giả sử có một bảng tính như sau:

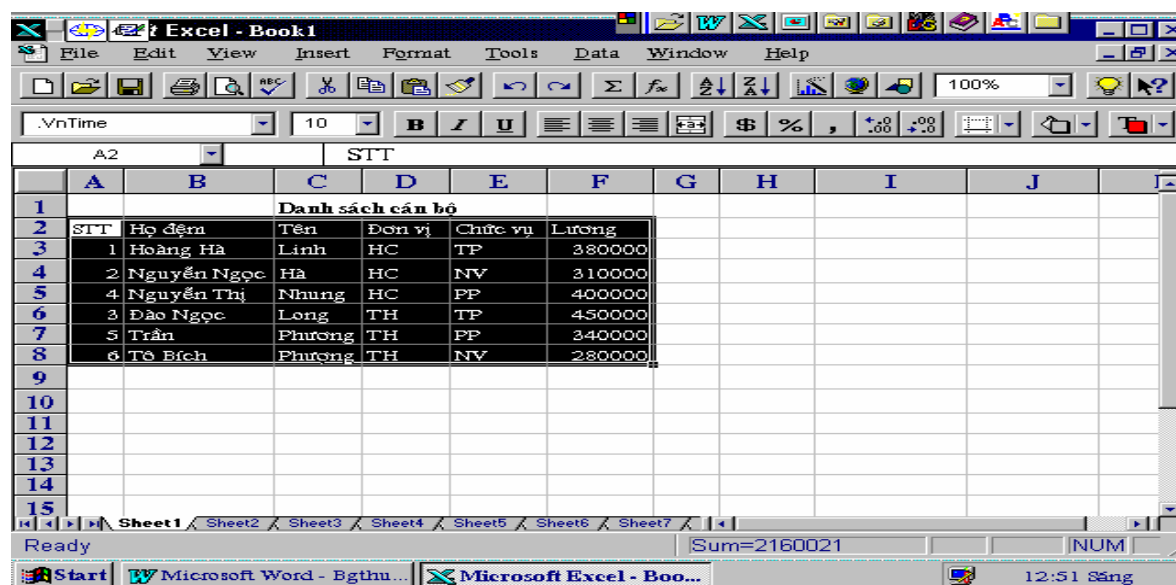
STT	Họ đệm	Tên	Đơn vị	Chức vụ	Lương
1	Hoàng Hà	Linh	HC	TP	380000
2	Nguyễn Ngọc	Hà	HC	NV	310000
3	Đào Ngọc	Long	TH	TP	450000
4	Nguyễn Thị	Nhung	HC	PP	400000
5	Trần	Phượng	TH	PP	340000
6	Tô Bích	Phượng	TH	NV	280000

Các bước thực hiện như sau:

* **Sắp xếp số liệu:** Trong bảng này, chúng ta cần tính tổng số lương được lĩnh của các cán bộ trong từng đơn vị, vì vậy khoá cần sắp xếp là đơn vị.

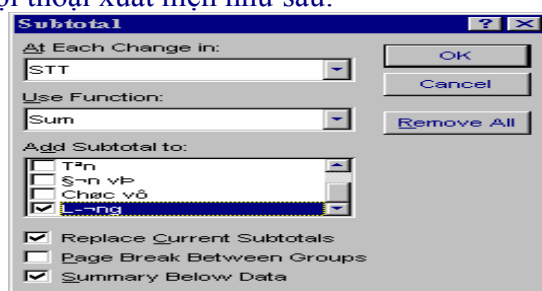
- Quét khối A2:F8/ vào bảng chọn Data/ Sort
- Trong hộp thoại Sort by, chọn Đơn vị, Ascending
- Trong mục My list has, chọn Header row
- Nhấn nút OK hoặc Enter

Chúng ta nhận được như sau:

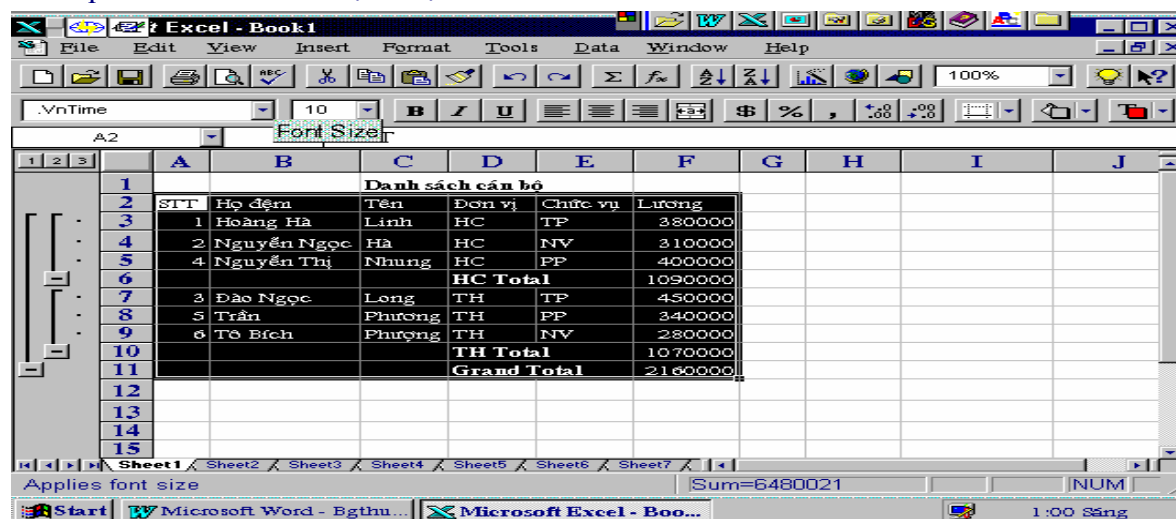


* Tính tổng nhóm

- Đưa con trỏ vào vùng CSDL
- Vào bảng chọn Data/Subtotals, hộp hội thoại xuất hiện như sau:
- At Each Change in: Chọn đơn vị (trường cần tạo nhóm tổng hợp).
- Use Function: Chọn Sum (hàm cần tính toán thống kê)
- Add Subtotal to: Chọn Lương (những trường cần tính toán).
- Nhấn nút OK hoặc Enter

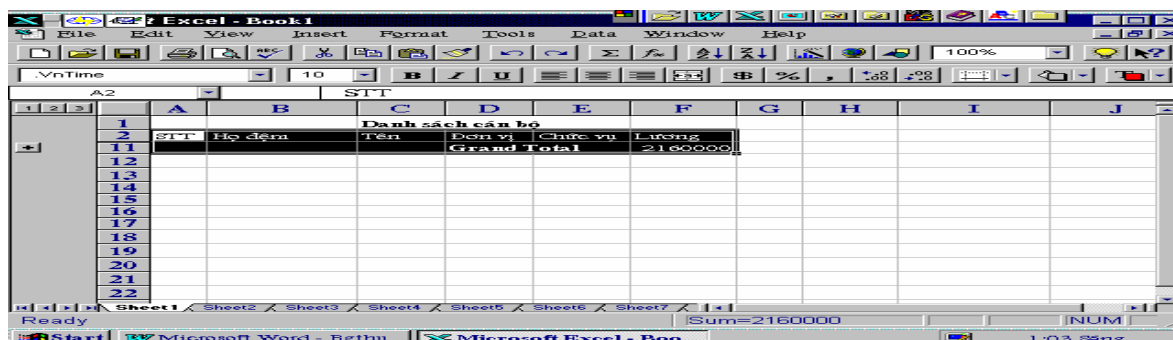


Kết quả trên màn hình nhận được như sau:

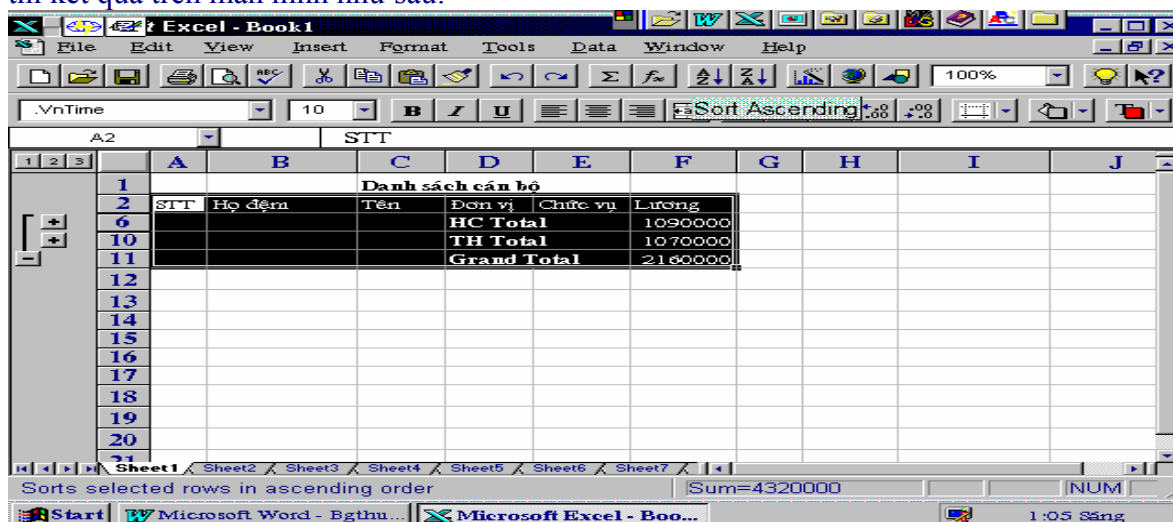


Bên trái bảng chọn xuất hiện các nút điều khiển thứ bậc, các nút có ý nghĩa như sau:

- Nút số 1: Chỉ hiện dòng Grand Total. Nếu chúng ta chọn nút này thì kết quả nhận được như sau:



Nút số 2: Chỉ hiện các dòng Toatal và dòng Grand Total. Nếu chúng ta chọn nút này thì kết quả trên màn hình như sau:



- Chọn nút số 3: Xuất về trạng thái kết quả ban đầu khi chưa chọn 1, chưa chọn 2.

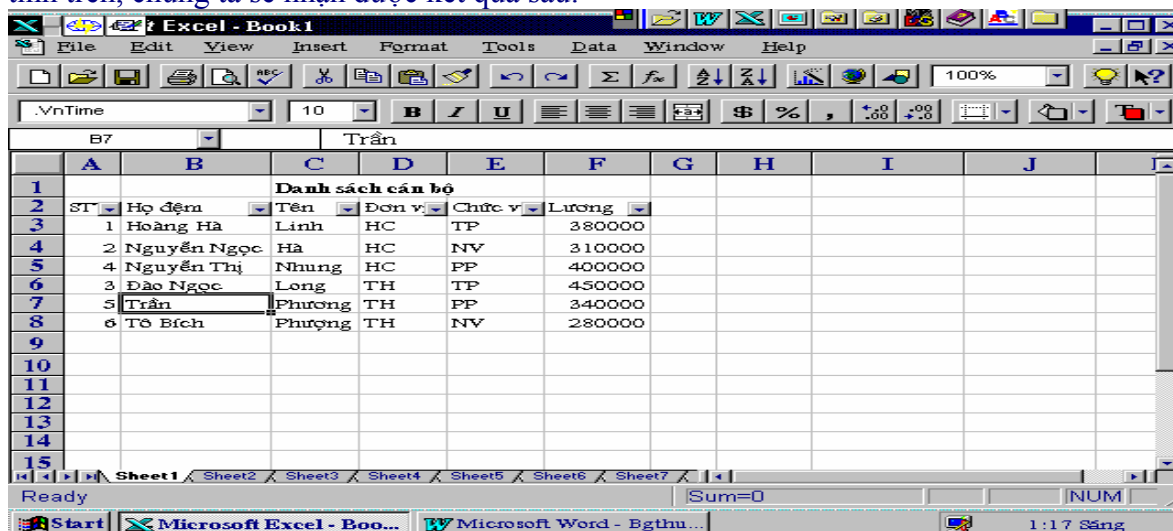
Nếu muốn huỷ bỏ việc tính tổng theo nhóm như trên để trở về dữ liệu bình thường thì chúng ta tiến hành như sau: Vào Data/ chọn Subtotals/ Remove all.

b) Sử dụng chức năng của AUTOFILTER

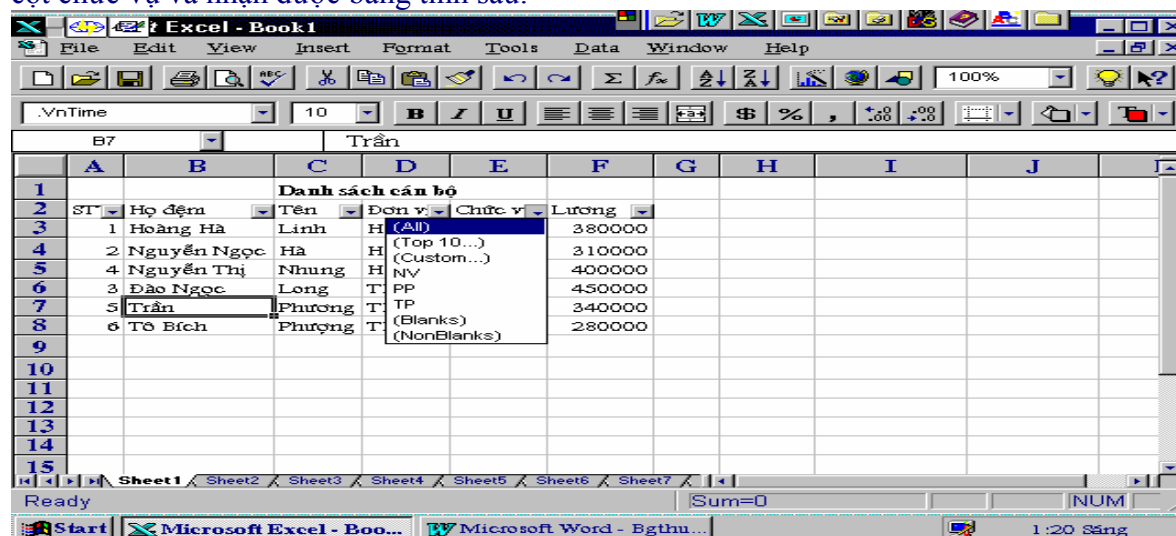
Chúng ta có thể sử dụng chức năng AutoFilter để đưa vào các tiêu chuẩn so sánh. Các bước tiến hành như sau:

- Tạo vùng dữ liệu/ Vào bảng chọn Data/ Filter/ AutoFilter.

Khi này, trên mỗi tiêu đề của cột, ở phía bên trái có mũi tên quay xuống. Với bảng tính trên, chúng ta sẽ nhận được kết quả sau:

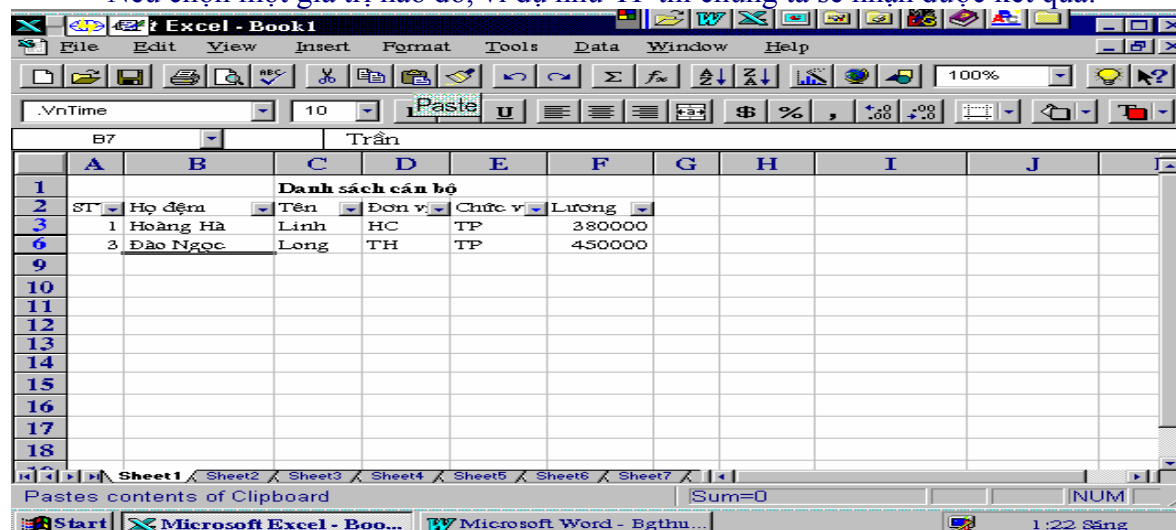


Nếu chọn tiêu chuẩn tìm kiếm theo chức vụ, chúng ta sẽ nhấn vào mũi tên bên phải cột chức vụ và nhận được bảng tính sau:



Nếu chọn All, tất cả các bản ghi thỏa mãn.

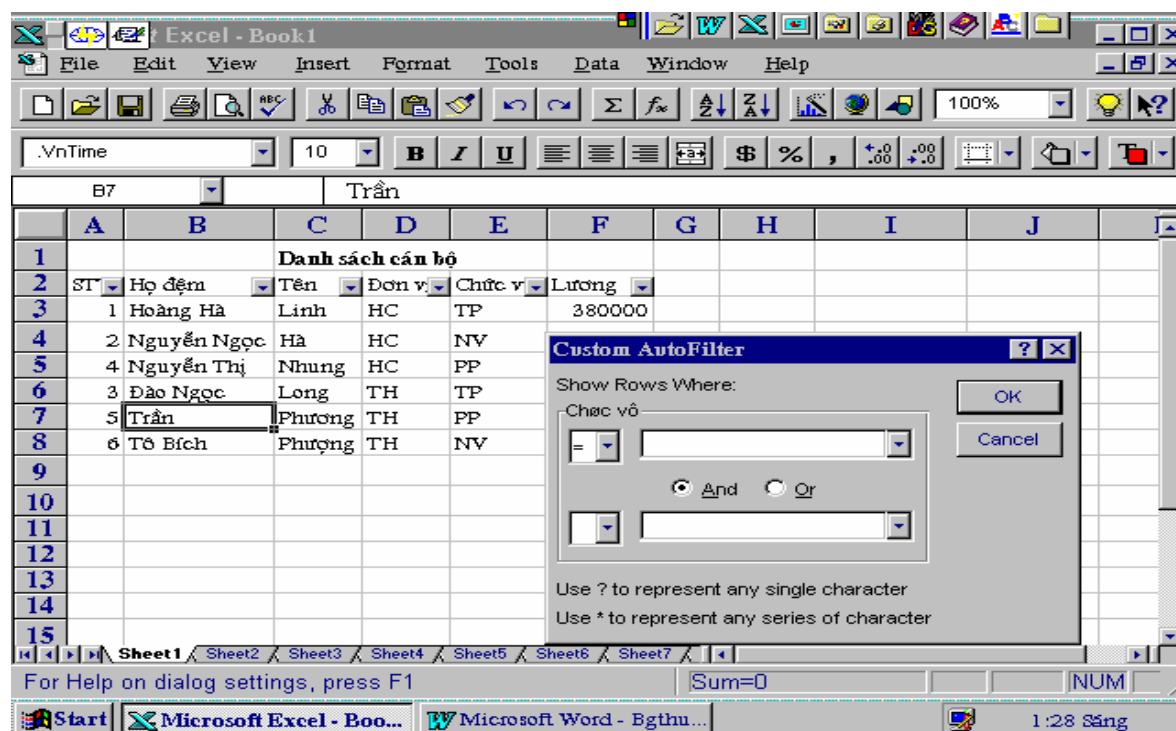
Nếu chọn một giá trị nào đó, ví dụ như TP thì chúng ta sẽ nhận được kết quả:



Nếu cần chọn tiêu chuẩn tìm kiếm những người có chức vụ nào thì chúng ta nhấn chuột vào dòng đó. Nếu muốn cho hiện tất cả mọi chức vụ thì nhấn chuột vào nút All.

Nếu muốn lựa chọn những người không có chức vụ nào (hay nói một cách tổng quát là cột này để trống) nhấn vào nút Blanks. Chọn những bản ghi mà trường này không trống, nhấn chuột vào NonBlanks.

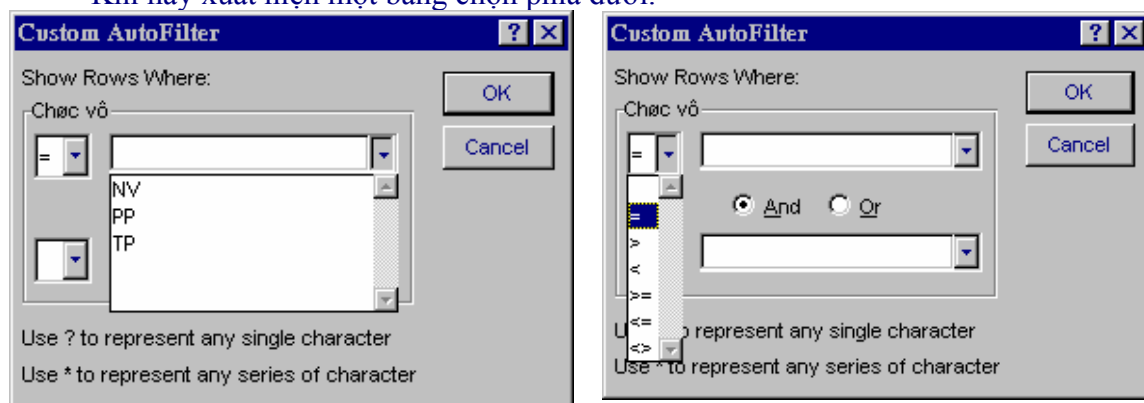
Bạn có thể nhấn chuột vào nút Custom.... để đưa các tiêu chuẩn theo ý của mình. Khi này, một hộp thoại xuất hiện. Chúng ta nhấn chuột vào mũi tên bên trái để chọn các khả năng so sánh:



Nếu chúng ta cần chọn bằng một giá trị nào đó, chúng ta có thể điền luôn vào ô trống phía trên. Trong trường hợp chúng ta muốn xem danh sách các loại giá trị có trong cột này, chúng ta nhấn chuột vào mũi tên bên phải. Khi này một danh sách các giá trị xuất hiện ở phía dưới, chúng ta chỉ việc nhấn chuột để chọn.

Nếu chúng ta cần chọn một giá trị theo dạng so sánh, chúng ta nhấn chuột ở mũi tên bên trái của cột.

Khi này xuất hiện một bảng chọn phía dưới.



Chú ý: dấu <> là dấu khác

Trong ví dụ trên, giả sử chúng ta chọn giá trị <> (khác) và đưa vào ô trống thứ nhất một giá trị, ví dụ như NV (nhân viên), chúng ta sẽ nhận được kết quả là các bản ghi không phải là NV (nhân viên).

ST	Họ đệm	Tên	Đơn vị	Chức vụ	Lương
1	Hoàng Hà	Linh	HC	TP	380000
4	Nguyễn Thị	Nhung	HC	PP	400000
3	Đào Ngọc	Long	TH	TP	450000
5	Trần	Phường	TH	PP	340000

4 of 6 records found

Bây giờ chúng ta kết hợp với một trường khác nữa, ví dụ như trường lương, chúng ta hãy chọn thêm điều kiện $Lương \geq 400000$, kết quả nhận được là:

ST	Họ đệm	Tên	Đơn vị	Chức vụ	Lương
4	Nguyễn Thị	Nhung	HC	PP	400000
3	Đào Ngọc	Long	TH	TP	450000

2 of 6 records found

Như vậy, chúng ta đã chọn được các bản ghi tiêu chuẩn sau:

(Chức vụ $\neq$) AND (Lương ≥ 400000)

Chúng ta nhấn tiếp mũi tên ở cột Lương, chọn Custom.... Khi này xuất hiện một hộp thoại như sau:

Custom AutoFilter

Show Rows Where:

L-ương

$\geq$ 400000

☒ And ☐ Or

Use ? to represent any single character
Use * to represent any series of character

OK Cancel

Chúng ta nhấn chuột vào nút **And** và sau đó nhấn nút chuột vào nút **♦** bên trái của dòng thứ hai và tương tự như trên, chọn $Lương \leq 430000$, chúng ta nhận được kết quả sau:

Tên		Danh sách cán bộ			Lương
STT	Họ đệm	Tên	Đơn vị	Chức vụ	Lương
4	Nguyễn Thị	Nhung	HC	PP	400000
3	Đào Ngọc	Long	TH	TP	450000
5	Trần	Phương	TH	PP	340000

Như chúng ta đã thấy, bản ghi số 3 không thỏa mãn. Như vậy, chúng ta đã chọn được các bản ghi theo tiêu chuẩn sau:

(Chức vụ $\in$ NV) AND (Lương ≥ 400000 AND Lương ≤ 430000)

Bây giờ chúng ta lại nhấn nút **♦** của cột Lương, chọn Custom.. thay nút **And** bằng việc nhấn nút **or**. Nhấn nút **♦** bên trái của dòng thứ hai và tương tự như trên chọn $Lương \leq 350000$, chúng ta nhận được giá trị như sau:

Tên		Danh sách cán bộ			Lương
STT	Họ đệm	Tên	Đơn vị	Chức vụ	Lương
4	Nguyễn Thị	Nhung	HC	PP	400000
3	Đào Ngọc	Long	TH	TP	450000
5	Trần	Phương	TH	PP	340000

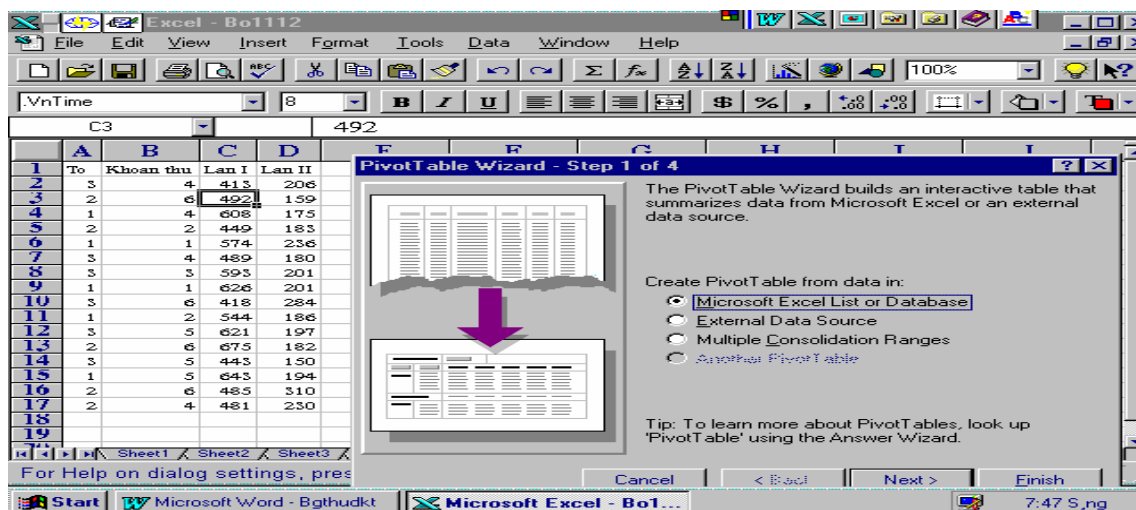
Trong trường hợp này, chúng ta đã chọn được các bản ghi theo tiêu chuẩn.

(Chức vụ $\in$ NV) AND (Lương ≥ 400000 OR Lương ≤ 350000)

c) Lệnh DATA/PIVOT TABLE

Từ EXCEL 5.0 trở đi trong Menu Data có thêm phần tổng kết Pivot Table (còn gọi là chuyển vị) là một công cụ mạnh mẽ để rút các thông tin cần thiết khi theo dõi hay tổng kết một bảng dữ liệu.

Giả sử có 1 bảng dữ liệu có nhiều cột, mỗi cột có tên ở hàng trên cùng và nhiều bản ghi (Record). Dùng chuột bôi đen cả bảng (hoặc có thể để con trỏ tại ô hiện thời thuộc vùng dữ liệu) sau đó vào Data/ Pivot Table, lần lượt có 4 bước đối thoại với máy:

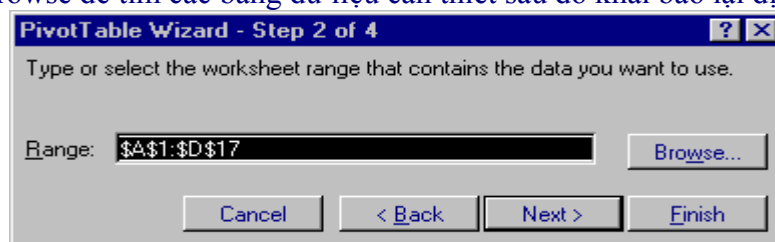


- **Bước 1:** Bảng Pivot table nêu 4 nguồn dữ liệu:

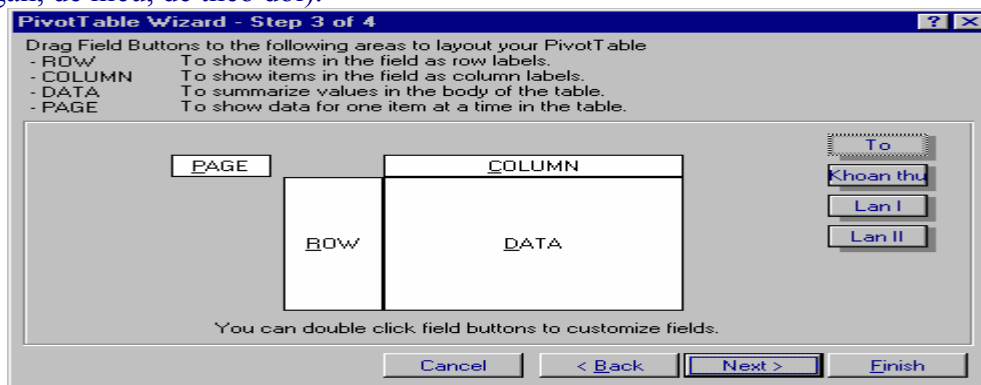
- + Microsoft Excel list ở Databáe: Danh sách hay bảng dữ liệu của Excel.
- + External Data source: Tập tin hay bảng dữ liệu từ các nguồn khác.
- + Multiple consolidation ranges: Nhiều dãy dữ liệu liên kết trong các bảng tính của Excel.
- + Another Pivot Table: Từ 1 bảng tổng kết khác đã có trong các bảng tính của Excel.

Thường ta để nguyên không thêm bớt gì (vì Excel đã được bôi đen nguồn dữ liệu I) sau đó chọn option Next để chuyển sang bước tiếp.

- **Bước 2:** Excel thông báo đại chỉ (dãy tham chiếu) của bảng dữ liệu mà ta đã chọn. Nếu không thấy có gì sai thì chọn bước sau bằng option Next, nếu cần thay đổi thì gõ lại địa chỉ hay dùng Browse để tìm các bảng dữ liệu cần thiết sau đó khai báo lại địa chỉ.



- **Bước 3:** Đây là bước quyết định nội dung bảng tổng kết. Ta có hộp thoại 2 vùng, vùng phải ghi tên các trường (tức là tên các cột của bảng dữ liệu, do đó khi chọn tên cột nên chọn ngắn, dễ hiểu, dễ theo dõi).

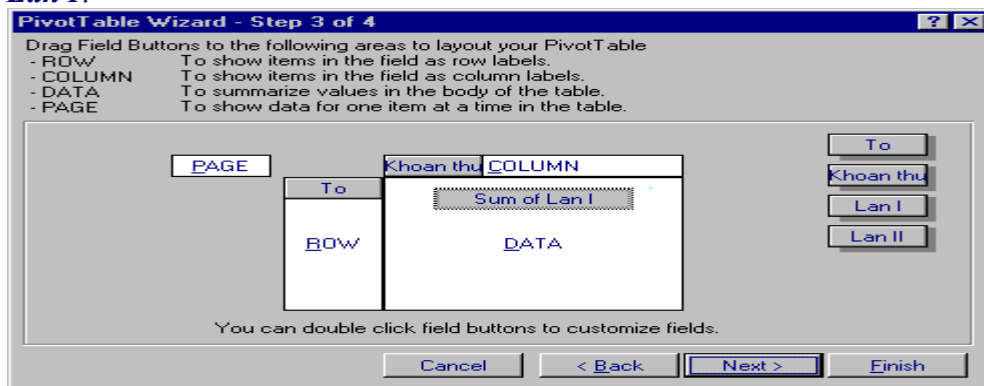


Vùng trái có 2 nửa, nửa trên ghi tóm tắt các nội dung của các từ Row (trường mà ta định đặt ở hàng), Column (trường mà ta định đặt ở cột), Data (trường số liệu mà ta định tổng kết), Page (trường mà ta định đặt ở trang).

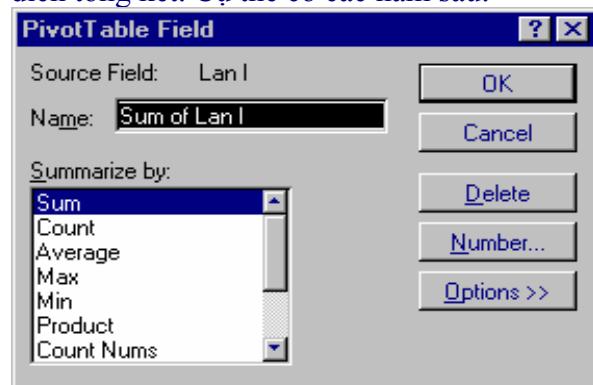
Dùng chuột nhấp 1 tên ở vùng phải, kéo và thả vào Page, sau đó nhấp một tên thả vào Row, nhấp 1 tên thả vào Column, nhấp một tên thả vào Data.

(Ta có thể nhấp nhiều tên thả vào Row, Column hay Data và một tên trường có thể xuất hiện nhiều lần ở Data với nhiều hàm khác nhau nhưng bước đầu không nên thả nhiều vì bảng tổng kết sẽ rối, khó nhìn, khó hiểu).

Khi thả tên vào Data ta thấy xuất hiện hàm SUM (thí dụ ta kéo trường Lan I thì có **Sum of Lan I**:

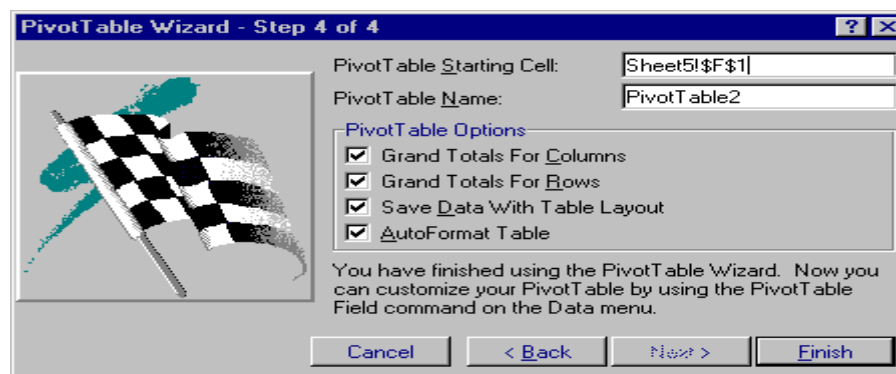


Ta có thể nhấp kép để hiện ra bảng các hàm có thể dùng để lựa chọn hàm thích hợp với mục đích tổng kết. Cụ thể có các hàm sau:



Sum: Tổng
 Count: Đếm số lần gặp
 Average: Số trung bình
 Max: Giá trị lớn nhất
 Min: Giá trị nhỏ nhất
 Product: Tích số
 Count Num: Số bản ghi
 Stdev: Độ lệch chuẩn của mẫu
 Stdevp: Độ lệch chuẩn tổng thể
 Var: Phương sai mẫu
 Varp: Phương sai tổng thể

Mỗi lần chọn xong chỉ tiêu thì vào OK



Sau khi thả hết các trường tính theo các chỉ tiêu theo mục đích lựa chọn ta vào Next của hộp đối thoại Bước 3 (Step 3) để chuyển sang bước 4.

- Bước 4: Chọn vị trí đặt bảng Pivot Table/ Finish, kết quả ta được bảng sau:

		Khoan thu						
To	Data	1	2	3	4	5	6	Grand Total
1	Average of Lan I	600	544	#DIV/0!	608	643	#DIV/0!	599
	Average of Lan II	218.5	186	#DIV/0!	175	194	#DIV/0!	198.4
2	Average of Lan I	#DIV/0!	449	#DIV/0!	481	#DIV/0!	550.6667	516.4

	Average of Lan II	#DIV/0!	183	#DIV/0!	230	#DIV/0!	217	212.8
3	Average of Lan I	#DIV/0!	#DIV/0!	593	451	532	418	496.167
	Average of Lan II	#DIV/0!	#DIV/0!	201	193	173.5	284	203
	Total Average of Lan I	600	496.5	593	498	569	517.5	534.625
	Total Average of Lan II	218.5	184.5	201	198	180.33	233.75	204.625

2.3. Lệnh giải bài toán quy hoạch tuyến tính (TOOLS/SOLVER)

a) Giới thiệu chung về lệnh Tools/Solver

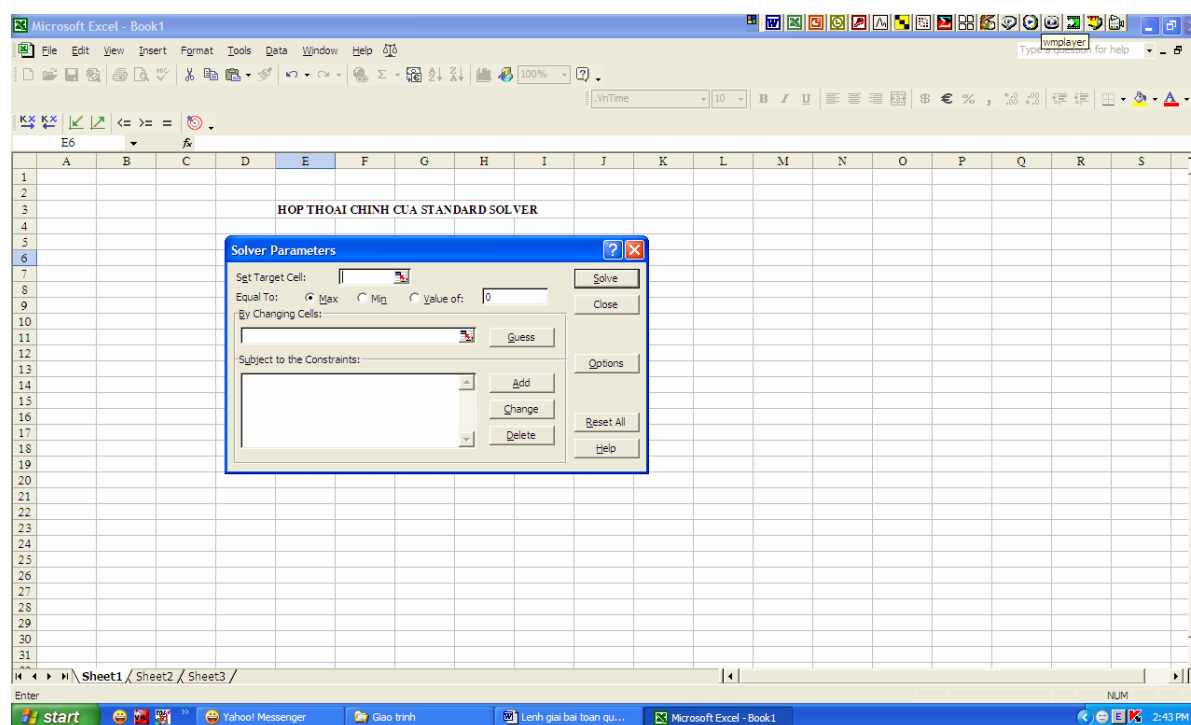
Khi cài đặt đầy đủ bộ Microsoft Office (từ Office 97 trở lên), Standard Solver (Solver chuẩn) nằm trong menu Tools của Microsoft Excel. Solver chuẩn có những công cụ cơ bản như thuật toán Gradient rút gọn tổng quát GRG (Generalized Reduced Gradient) để tối ưu hoá các bài toán phi tuyến do Leon Lasdon, University of Texas at Austin và Allan Waren, Cleveland State University phát triển. Đối với các bài toán tuyến tính và biến nguyên (integer variable) Solver dùng phương pháp đơn hình (simple method) và phương pháp nhánh và cận (branch and bound method) do John Watson và Dan Fylsstra, thuộc Frontline Systems, Inc thực hiện.

Standard Solver gồm 2 công cụ:

- Standard GRG Nonlinear dùng thuật toán Generalized Reduced Gradient (GRG) để giải các bài toán quy hoạch phi tuyến tròn tru với quy mô 200 biến quyết định (decision variables, để đơn giản, sau này gọi là biến) và 100 ràng buộc (constraints) cộng với 400 ràng buộc cận đặt trên biến (Bounds on Variables).
- Standard Simplex Linear dùng thuật toán đơn hình (Simplex) để giải các bài toán quy hoạch tuyến tính với quy mô 400 biến và 200 ràng buộc cộng với 800 ràng buộc cận đặt trên biến.

Những khái niệm cơ bản trong Solver chuẩn:

- Hộp thoại “Solver Parameters” là giao diện chính để làm việc với Standard Solver. Để hiển thị hộp thoại này, trong môi trường Excel, click menu “Tools”, chọn Solver. Hình sau đây sẽ hiển thị hộp thoại “Solver Parameters”:



Mục đích chính của Solver là tìm nghiệm nghĩa là các giá trị cho các biến trong mô hình (Solver đặt tên là “Changing Cells”). Các giá trị này thỏa mãn các ràng buộc (constraints) và làm cực đại hoặc cực tiểu mục tiêu (objective) (Solver đặt tên là “Set Target Cell”).

- Biến số và tham số:

+ Một số giá trị đầu vào có thể là con số cố định ta không thể thay đổi trong quá trình tìm nghiệm, chẳng hạn như lãi suất, giá sản phẩm, v.v..., ta gọi những giá trị này là tham số của mô hình. Dĩ nhiên các tham số này sẽ thay đổi trong từng phương án, từng kịch bản.

+ Những giá trị đầu vào khác có thể là con số thay đổi hoặc ta muốn điều khiển thay đổi trong quá trình tìm nghiệm gọi là các biến, hay biến quyết định hay “Changing Cells”. Solver sẽ tìm giá trị tối ưu cho những biến này.

- Hàm mục tiêu (Objective Function): Con số nằm trong ô ta muốn cực đại hoặc cực tiểu được gọi là hàm mục tiêu hoặc Set Target Cell. Ô này sẽ được chỉ định trong khung Set target Cell của hộp thoại “Solver Parameters”. Hàm mục tiêu có thể là lợi nhuận nhiều nhất (tìm cực đại), chi phí ít nhất (tìm cực tiểu).

Có thể có mô hình Solver không có hàm mục tiêu để tìm cực trị. Khi đó khung Set Target Cell trong hộp thoại “Solver parameters” để trống. Trong trường hợp này Solver chỉ đơn giản tìm nghiệm thỏa mãn các ràng buộc. Nói cách khác, đây là việc tìm nghiệm hệ phương trình (hệ phương trình ràng buộc) khi biết giá trị vế phải của hệ ràng buộc.

Solver cũng có thể tìm giá trị biến quyết định khi cho trước hàm mục tiêu. Thực ra việc đưa giá trị xác định vào khung Set Target Cell của hộp thoại Solver cũng tương đương để trống khung này và đưa vào ràng buộc cân bằng trong khung danh mục ràng buộc Subject to the Constraint của hộp thoại “Solver Parameters”. Nếu bài toán của ta chỉ đơn lẻ có một giá trị Set Target Cell và một biến quyết định thì lúc đó chẳng cần Solver làm gì mà dùng lệnh Goal Seek trong menu Tools.

- Ràng buộc (Constraints) trong Excel Solver được định rõ bằng cách cho ô hoặc các ô tham chiếu ở vế trái ví dụ A1 hoặc A1:A5, một dấu quan hệ ($\leq$, $=$, hoặc $\geq$) và một biểu thức ở vế bên phải. Mặc dù Excel Solver cho phép đưa bất kỳ biểu thức số học vào vế phải nhưng ta chỉ nên đưa vào đó hằng số hoặc tham chiếu đến ô có giá trị hằng số. Một giá trị hằng số đối với Solver là một giá trị không phụ thuộc vào bất cứ biến quyết định nào.

Một ràng buộc như $A1:A5 \leq 10$ là viết gọn cho $A1 \leq 10$, $A2 \leq 10$, $A3 \leq 10$, $A4 \leq 10$, $A5 \leq 10$. Một ràng buộc $A1:A5 \leq B1:B5$ là viết gọn cho $A1 \leq B1$, $A2 \leq B2$, $A3 \leq B3$, $A4 \leq B4$, $A5 \leq B5$.

Một ràng buộc khác là $A1:A5 = \text{integer}$, trong đó A1:A5 là biến quyết định. Điều này quy định rõ ràng các giá trị nghiệm cho A1 đến A5 phải là số nguyên chẳng hạn -1, 0 hoặc 2 nằm trong phạm vi sai số nhỏ. Ràng buộc tương tự như $A1:A5 = \text{binary}$ cho biết giá trị nghiệm là số nhị phân (0, 1).

- Phân loại nghiệm: Khả thi (Feasible) và tối ưu (Optimal)

Một nghiệm (tập các giá trị cho biến quyết định) mà với nghiệm này, tất cả các ràng buộc trong mô hình Solver thỏa mãn thì được gọi là nghiệm khả thi. Trong một số bài toán, một nghiệm khả thi là biết trước. Trong một số bài toán khác, việc tìm nghiệm khả thi có thể là phần khó nhất của bài toán.

Một nghiệm tối ưu là một nghiệm khả thi mà với nghiệm đó hàm mục tiêu đạt đến giá trị cực đại (hoặc cực tiểu). Một nghiệm tối ưu toàn cục (globally optimal solution) là một nghiệm mà với nghiệm đó không có bất kỳ nghiệm khả thi nào khác có giá trị hàm mục tiêu tốt hơn. Một nghiệm tối ưu cục bộ (locally optimal solution) là một nghiệm mà với nghiệm đó không có bất kỳ nghiệm khả thi nào ở “lân cận” có giá trị hàm mục tiêu tốt hơn.

Solver được thiết kế để tìm nghiệm khả thi và nghiệm tối ưu. Trong trường hợp tốt nhất nó sẽ tìm được nghiệm tối ưu toàn cục nhưng điều này không phải lúc nào cũng làm được. Nói chung nó sẽ tìm nghiệm tối ưu cục bộ và sẽ dừng lại sau một thời gian nào đó với

nghiệm tốt nhất nó đã tìm được cho đến lúc đó. Điều quan trọng nhất đối với ta là tìm một nghiệm tốt nhất tức là tìm một nghiệm tốt hơn nghiệm hoặc đám đông nghiệm ta dùng sử dụng hiện nay.

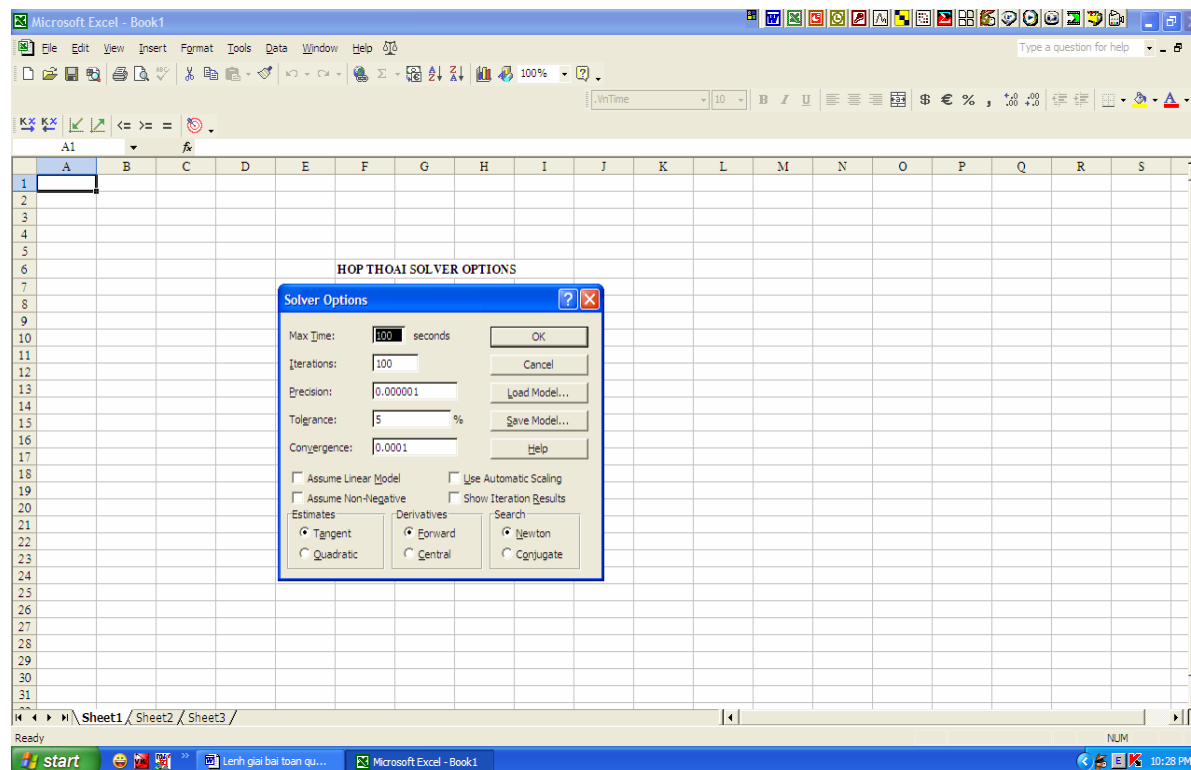
Solver có thể tìm được một nghiệm tối ưu toàn cục, tối ưu cục bộ hay nghiệm tốt hay không sẽ tùy thuộc bản chất quan hệ toán học giữa các biến, hàm mục tiêu và các ràng buộc và vào thuật toán học được dùng để tìm nghiệm.

Các công cụ trong Solver chuẩn:

- Đơn hình tuyến tính Simplex Solver dùng để tìm nghiệm tối ưu cho các bài toán mà hàm mục tiêu và ràng buộc là hàm tuyến tính của biến. Đối với bài toán này, Solver thường có thể tìm được nghiệm tối ưu toàn cục (nếu nó tồn tại). Vì hàm tuyến tính luôn có thể tăng hoặc giảm không bị giới hạn nên nghiệm tối ưu luôn xác định bởi các ràng buộc.
- Quy hoạch phi tuyến GRG Solver (nonlinear GRG Solver) tìm nghiệm tối ưu cho các bài toán mà hàm mục tiêu và các ràng buộc đều là hàm trơn tru của các biến. Đối với bài toán này, Solver thường có thể tìm được một nghiệm tối ưu cục bộ, nếu nó tồn tại, nhưng nghiệm này có thể hoặc không phải là nghiệm tối ưu toàn cục. Một hàm mục tiêu phi tuyến có thể có một đỉnh (peak) hoặc một lõm (valley) tự nhiên nhưng trong phần lớn các bài toán nghiệm tối ưu được xác định một phần hoặc toàn bộ qua các ràng buộc.

Trong Solver chuẩn, phi tuyến GRG Solver là công cụ mặc định, nghĩa là nếu không chỉ rõ giải bài toán bằng công cụ nào (đơn hình hay GRG) thì đương nhiên Solver sẽ dùng công cụ GRG để giải ngay cả khi bài toán là tuyến tính. Công cụ RGR có thể được dùng cho các bài toán với các hàm hoàn toàn tuyến tính nhưng không hiệu quả bằng công cụ đơn hình Simplex Solver.

Để chỉ định dùng công cụ đơn hình Simplex cho các bài toán tuyến tính, từ hộp thoại Solver Parameter, dùng chuột click vào khung “Options” sẽ xuất hiện hộp thoại “Solver Options” như sau:



Đánh dấu vào khung “Assume Linear Models” của hộp này, Solver chuẩn sẽ dùng thuật giải đơn hình Simplex để giải các bài toán tuyến tính.

b) Bài toán quy hoạch tuyến tính

Bài toán quy hoạch tuyến tính đã được sử dụng rộng rãi mấy chục năm nay như giải các bài toán trong sản xuất và lập kế hoạch. Trong nông nghiệp nhiều bài toán đã trở thành những thí dụ kinh điển về ứng dụng quy hoạch tuyến tính như việc lập kế hoạch sản xuất, xác định khẩu phần thức ăn hoặc ứng dụng các phương pháp riêng của quy hoạch tuyến tính như các bài toán vận tải, bài toán phân phối.

Dạng tổng quát của một mô hình bài toán quy hoạch tuyến tính:

- Hàm mục tiêu. $Z = \sum_{j=1}^n C_j X_j \rightarrow \text{Max (Min)}$

- Các ràng buộc làm cho hạn chế mục tiêu.

$$\sum_{j=1}^n a_{ij} X_j \geq A_j \quad (\text{Min}) \qquad \sum_{j=1}^n a_{ij} X_j \leq A_j \quad (\text{Max})$$

Với $X_j \geq 0 \quad (j = \overline{1, n}) \rightarrow$ Điều kiện đương nhiên

C_j, a_{ij}, A_j là các thông số, hằng số

X_j là biến số bậc nhất (có số mũ bằng 1)

Các bước lập mô hình bài toán quy hoạch tuyến tính trên EXCEL:

- Sắp xếp dữ liệu câu mô hình trên bảng tính: Dữ liệu của mô hình gồm các hệ số trong hàm mục tiêu, các hệ số trong hệ ràng buộc và giá trị vế phải của hệ ràng buộc. Có nhiều cách sắp xếp dữ liệu cho một bài toán cụ thể, nhưng ta nên theo một nguyên tắc chung. Đó là sắp xếp dữ liệu sao cho có ý nghĩa và mục đích càng rõ ràng càng tốt. Nên xem bảng tính như là một bản báo cáo thể hiện rõ các yếu tố quan trọng của bài toán dạng giải. Nên đặt các nhãn mô tả (labels) trong bảng tính để nhận biết các phần tử khác nhau của dữ liệu. Lưu ý rằng có thể một vài hoặc tất cả các hệ số và các giá trị của mô hình bài toán quy hoạch tuyến tính được tính từ các dữ liệu khác (dữ liệu sơ cấp). Trong trường hợp này nên đưa các dữ liệu sơ cấp vào bảng tính và dùng các công thức thích hợp để tính các hệ số và các giá trị cần thiết cho việc hình thành mô hình quy hoạch tuyến tính. Như vậy, khi các dữ liệu sơ cấp thay đổi thì các hệ số và các giá trị của mô hình quy hoạch tuyến tính sẽ tự động thay đổi.

- Dành các ô riêng biệt trong bảng tính để thực hiện từng biến quyết định trong mô hình. Mặc dù ta có thể dùng bất kỳ ô trống nào (empty cell) trong bảng tính để thực hiện biến quyết định nhưng có lẽ cách sắp xếp tốt nhất là để các ô thể hiện biến quyết định song song với cấu trúc dữ liệu. Điều này sẽ hữu ích trong việc thiết lập các công thức cho hàm mục tiêu và các ràng buộc. Nếu có thể đặt các ô thể hiện biến quyết định trong cùng một vùng câu bảng tính. Nên dùng các nhãn mô tả để nhận biết rõ ràng ý nghĩa của các ô này.

- Tạo công thức trong một ô bảng tính tương ứng với hàm mục tiêu trong mô hình. Công thức tương ứng với hàm mục tiêu được tạo ra bằng cách tham chiếu đến các ô dữ liệu, ở đó các hệ số của hàm mục tiêu đã được nhập vào (hoặc được tính) và tham chiếu đến các ô tương ứng thể hiện các biến quyết định.

- Tạo công thức vế trái của các ràng buộc. Với mỗi ràng buộc, tạo một công thức trong một ô riêng biệt trong bảng tính, ô này tương ứng với vế trái câu ràng buộc. Công thức tương ứng với vế trái của ràng buộc được tạo ra bằng cách tham chiếu đến các ô dữ liệu, ở đó các hệ số cho các ràng buộc này đã được nhập vào (hoặc đã được tính) và tham chiếu đến các biến quyết định thích hợp. Có thể nhiều công thức ràng buộc có cấu trúc giống nhau, vì thế, khi cho phép ta nên tạo ra công thức ràng buộc có thể được copy để dùng cho các công thức ràng buộc khác. Điều này không những giảm bớt công sức thiết lập mô hình bảng tính mà còn hạn chế những sai sót khi biên tập bảng tính. Cũng nên dùng tô bóng (shading), màu nền (background color) và khung viền (border) để nhận rõ các ô thể hiện biến quyết định, ràng buộc và hàm mục tiêu trong mô hình bảng tính. Điều này cho phép người sử dụng bảng tính phân biệt rõ ràng giữa đâu là dữ liệu có thể thay đổi với các phần tử khác của mô hình.

- Làm việc với Solver nghĩa là đưa mô hình bảng tính vừa thiết lập vào khung hội thoại Solver. Ta dùng hộp thoại “Solver Parameters” để chỉ ra cho Solver (i) những ô dành cho biến quyết định, (ii) ô để tính giá trị hàm mục tiêu, (iii) những ô tính giá trị về trái của các ràng buộc và (iv) những ô chứa giá trị về phải cùng các dấu thể hiện tương ứng giữa về trái và về phải của các ràng buộc. Với các biến quyết định có ràng buộc cận trên (cận dưới) cũng đưa vào hộp thoại này.

Vì ở đây đang đề cập đến các bài toán quy hoạch tuyến tính nên ta chỉ định dùng công cụ đơn hình Simplex trong Solver cho các bài toán tuyến tính. Để làm việc này, từ hộp thoại Solver Parameters click vào khung Options, đánh dấu vào khung Assume Linear Models. Nếu các biến quyết định có ràng buộc không âm thì ta sẽ đánh dấu vào khung Assume Non-negative.

Sau khi đã hoàn thành các bước trên, công việc còn lại rất đơn giản, đó là lệnh cho Solver giải bài toán ta đã thiết lập trong bảng tính EXCEL bằng cách nhấp chuột vào khung Solver nằm ở trên cùng, phía phải của hộp thoại Solver Parameters. Sau vài giây thì kết quả bài toán sẽ được hiện ra màn hình.

Trên đây là những hướng dẫn tổng quát. Kỹ năng thiết lập mô hình bảng tính trong EXCEL để giải các bài toán kinh tế kỹ thuật nói chung đòi hỏi thời gian tích lũy kinh nghiệm. Sau đây là ví dụ minh họa bài toán lập kế hoạch sản xuất:

Một trang trại cần lập kế hoạch sản xuất trồng trọt tối ưu trên diện tích hiện có. Vấn đề đặt ra là nên trồng bao nhiêu tấn điều và bao nhiêu tấn hồ tiêu để có lợi nhuận lớn nhất trong điều kiện hạn chế về đất, nước, con người. Các số liệu cụ thể về diện tích đất, nước, nhân công để sản xuất và khả năng tối đa của mỗi yếu tố được cho trong bảng sau:

Chỉ tiêu	Điều	Hồ tiêu	Khả năng nguồn lực huy động
Diện tích trên 1 tấn sản phẩm (ha)	2	3	50
Lượng nước trên 1 tấn sản phẩm ($10^3 m^3$)	6	4	90
Nhân công trên 1 tấn sản phẩm (công)	20	5	250
Lợi nhuận trên 1 tấn sản phẩm (100USD)	18	21	

Thiết lập mô hình toán:

- Các biến quyết định: Gọi x_1 và x_2 là số tấn điều và hồ tiêu cần được sản xuất.

- Hàm mục tiêu: Lập kế hoạch sản xuất theo mục tiêu cực đại lợi nhuận: $18x_1 - 21x_2 \rightarrow \text{Max}$

- Các ràng buộc:

+ Ràng buộc về diện tích: $2x_1 + 3x_2 \leq 50$

+ Ràng buộc về lượng nước: $6x_1 + 4x_2 \leq 90$

+ Ràng buộc về nhân lực: $20x_1 + 5x_2 \leq 250$

+ Ràng buộc đương nhiên: $x_1, x_2 \geq 0$

Mô hình hoá bài toán trên bảng tính EXCEL:

	A	B	C	D	E
1	Chỉ tiêu	Điều	Hồ tiêu	Tổng cộng	Khả năng max
2	Diện tích trên 1 tấn sản phẩm (ha)	2	3	=B2*B6+C2*C6	50
3	Lượng nước trên 1 tấn sản phẩm ($10^3 m^3$)	6	4	=B3*B6+C3*C6	90
4	Nhân công trên 1 tấn sản phẩm (công)	20	5	=B4*B6+C4*C6	250
5	Lợi nhuận trên 1 tấn sản phẩm (100USD)	18	21	=B5*B6+C5*C6	
6	Lượng sản xuất	X1	X2		

Ghi chú: các công thức ở D2, D3, D4, D5 là “giống nhau”, bạn có thể cố định địa chỉ cho B6 và C6 rồi copy cho 3 ô còn lại.

Khai báo mô hình trên Solver: Chọn menu Insert/Name/Define để đặt tên cho:

\$D\$5 là Tong_loi_nhuan

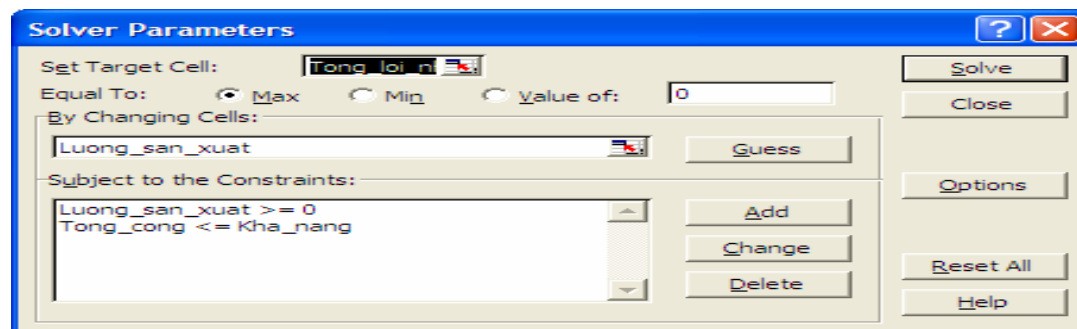
\$B\$6:\$C\$6 là Luong_san_xuat

\$D\$2:\$D\$4 là Tong_cong

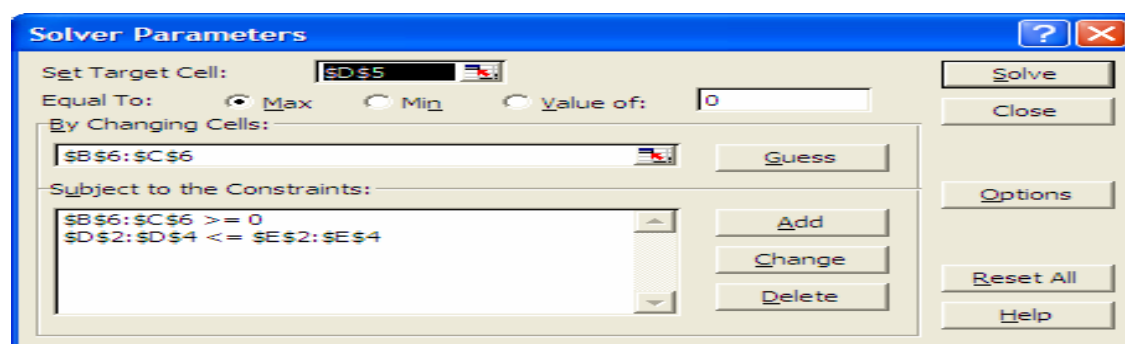
\$E\$2:\$E\$4 là Kha_nang

Ta có thể khai báo mô hình trên Solver như sau:

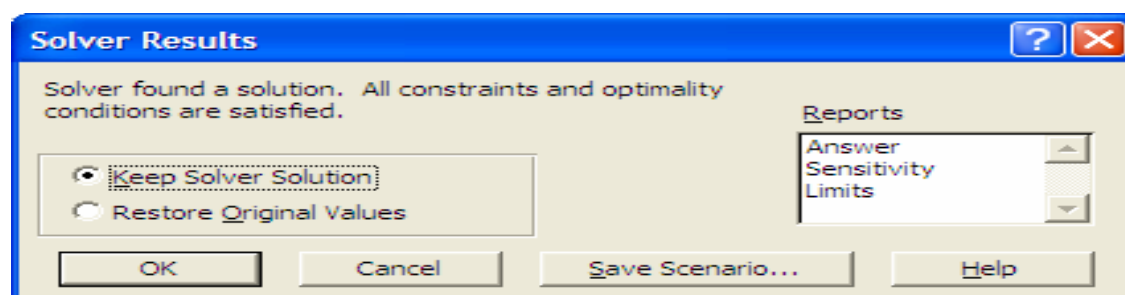
- Nhập tham số vào Solver bằng tên:



- Nhập tham số vào Solver bằng địa chỉ:



- Sau khi đã chuẩn bị xong, chọn Solver. Nếu kết quả tốt, EXCEL thông báo là “found a solution”:



Hãy chọn Keep Solver Solution để lưu kết quả trên bảng tính (nếu chọn Restore Original Values sẽ hủy kết quả Solver vừa tìm được và trả lại giá trị khởi động của các biến. Bạn có thể chọn Save Scenario để lưu kết quả mà Solver vừa tìm được mà không ảnh hưởng đến bảng tính gốc.

Ý nghĩa kết quả đối với nhà quản lý:

- Quan sát các ô kết quả:

	A	B	C	D	E
1	Chỉ tiêu	Điều	Hồ tiêu	Tổng cộng	Khả năng max
2	Diện tích trên 1 tấn sản phẩm (ha)	2	3	50	50
3	Lượng nước trên 1 tấn sản phẩm (10 ³ m ³)	6	4	90	90
4	Nhân công trên 1 tấn sản phẩm (công)	20	5	200	250
5	Lợi nhuận trên 1 tấn sản phẩm (100USD)	18	21	378	
6	Lượng sản xuất	7	12		

B6:C6 là biến quyết định và D5 là hàm mục tiêu.

- Phân tích kết quả của EXCEL:

Bạn có thể chọn một trong 3 báo cáo: Answer, Sensitivity, Limits. Khi phân tích các báo cáo này, người ta gọi là phân tích sau tối ưu.

+ Báo cáo Answer:

Target Cell (Max)

Cell	Name	Original Value	Final Value
\$D\$5	Tong_loi_nhuan	#VALUE!	378

Adjustable Cells

Cell	Name	Original Value	Final Value
\$B\$6	Lượng sản xuất điều	0	7
\$C\$6	Lượng sản xuất hồ tiêu	0	12

Constraints

Cell	Name	Cell Value	Formula	Status	Slack
\$D\$2	Diện tích trên 1 tấn sản phẩm (ha)	50	\$D\$2<=\$E\$2	Binding	0
\$D\$3	Lượng nước trên 1 tấn sản phẩm (10 ³ m3)g	90	\$D\$3<=\$E\$3	Binding	0
\$D\$4	Nhân công trên 1 tấn sản phẩm (công)	200	\$D\$4<=\$E\$4	Not Binding	50
\$B\$6	Lượng sản xuất điều	7	\$B\$6>=0	Not Binding	7
\$C\$6	Lượng sản xuất hồ tiêu	12	\$C\$6>=0	Not Binding	12

+ Báo cáo Sensitivity:

Adjustable Cells

Cell	Name	Final Value	Reduced Cost	Objective Coefficient	Allowable Increase	Allowable Decrease
\$B\$6	Lượng sản xuất điều	7	0	18	13.5	4
\$C\$6	Lượng sản xuất hồ tiêu	12	0	21	6	9

Constraints

Cell	Name	Final Value	Shadow Price	Constraint R.H. Side	Allowable Increase	Allowable Decrease
\$D\$2	Diện tích trên 1 tấn sản phẩm (ha)	50	5.4	50	17.5	10
\$D\$3	Lượng nước trên 1 tấn sản phẩm (10 ³ m3)	90	1.2	90	10	23.3
\$D\$4	Nhân công trên 1 tấn sản phẩm (công)	200	0	250	1E+30	50

+ Báo cáo Limits:

Cell	Name	Target Value
\$D\$5	Tong_loi_nhuan	378

Cell	Name	Value	Lower Limit	Target Result	Upper Limit	Target Result
\$B\$6	Lượng sản xuất điều	7	0	252	7	378
\$C\$6	Lượng sản xuất hồ tiêu	12	0	126	12	378

+ Thuật ngữ “Binding” để chỉ một ràng buộc có tính chất “bộc” với nghiệm tối ưu hay nó chứa nghiệm tối ưu. Các yếu tố thuộc về các ràng buộc “binding” sẽ trở thành các yếu tố khan hiếm, do đó, sẽ phát sinh khái niệm giá kinh tế hay giá mờ (Shadow Price).

+ Shadow Price là giá bóng phản ánh giá trị của việc tăng thêm một đơn vị nguồn lực bị khan hiếm sẽ làm cho hàm mục tiêu tăng (giảm) hoặc không tăng (giảm) bằng số đơn vị giá bóng. Trong bài toán này thì yếu tố diện tích và lượng nước trở nên khan hiếm và giá trị kinh tế của nó lần lượt là \$540/đơn vị và \$120/đơn vị. Ngược lại yếu tố nhân công không là yếu tố khan hiếm, số lượng nhân công có là 250, dư 50 (slack) so với số lượng cần là 200.

3. Ứng dụng một số chức năng của EXCEL trong quản trị kinh doanh

3.1. Lệnh Tools/Goal Seek

Bài toán: Phòng kinh doanh A dự kiến thuê 1 nhân viên mới. Mức lương bình quân chung 1 nhân viên của Phòng trước khi tuyển nhân viên mới là 714,444 ngàn đồng/ tháng. Ông Trưởng Phòng kinh doanh đề nghị với nhân viên mới 1 mức lương nhất định để làm cho mức lương bình quân chung 1 nhân viên của Phòng sau khi có thêm nhân viên mới là 700 ngàn đồng/ tháng nhưng không làm thay đổi mức lương của những người đã làm trước khi tuyển nhân viên mới này. Bằng kiến thức kinh tế anh (chị) hãy giúp ông Trưởng Phòng tính mức lương cho nhân viên mới.

Với bài toán này chúng ta dùng Hàm AVERAGE và dùng lệnh Tools/Goal Seek để giải bài toán trên.

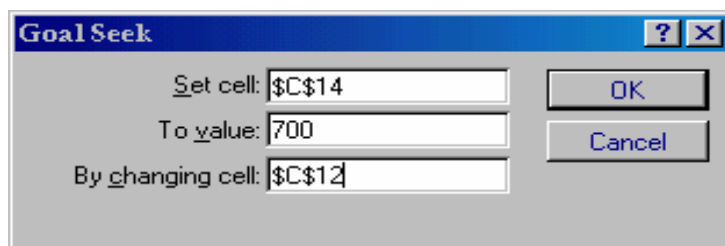
- Nhập số liệu và tính bình quân lương cho công nhân viên trước khi định thuê nhân viên mới:

The screenshot shows an Excel spreadsheet titled "Excel - Thqtcd2". The formula bar displays the formula `=AVERAGE(C3:C12)` in cell C14. The spreadsheet contains the following data:

STT	Họ Tên	Mức lương (1000 đ)
1	Lê Minh	780
2	Vũ Ngọc Bảo	750
3	Trịnh Phương Thanh	780
4	Lê Hồng Ngát	700
5	Trương Đức Thành	650
6	Lê Thị Thu Hà	600
7	Nguyễn Hồng Minh	780
8	Trần thị Thanh Lam	600
9	Phạm Thị Minh Thắng	790
10	Tô Cẩm Nhung	
	Bình quân	714.4444444

The status bar at the bottom shows "Page 66", "Sec 1", "34/34", "At 11.9cm Ln 17 Col 1", and "Sum=714.4444444".

Dùng lệnh Tools/Goal Seek để giải bài toán:



Chọn nút OK để có kết quả

Chú ý: Khi dùng hàm Average để tính trung bình lương cho những người đang làm cho phòng kinh doanh mặc dù người dự kiến tuyển vào chưa có tên trong danh sách lương nhưng ta vẫn tính vào hàm Average này vì nó liên quan đến việc hạch toán lương cho nhân viên mới thông qua lệnh Tools/Goal Seek.

Kết quả bài toán cho thấy rằng mặc dù lương của nhân viên cũ không thay đổi, tuyển nhân viên mới vào với mức lương thấp (họ chấp nhận được) đã làm cho mức lương bình quân chung sau khi tuyển thêm nhân viên vào giảm, chứng tỏ Phòng kinh doanh có hiệu quả. Như vậy Goal Seek có tác dụng làm thay đổi giá trị trong 1 ô để quy kết các giá trị đích.

3.2 Lệnh Data/Table

Ví dụ có 1 khoản tiền vay là \$125000, với lãi suất 7,5% và thời gian là 30 năm. Tính khoản tiền trả hàng tháng. Với mức lãi suất lần lượt là: 6,5%; 6,75%; 7%; 7,25%; 7,5%; 7,75%; 8%; 8,25%; 8,5%. Khoản tiền vay cũng lần lượt tương ứng là \$115000; \$120000; \$130000; \$135000; \$140000 thì tiền trả hàng tháng là bao nhiêu với thời gian cho vay là 30 năm.

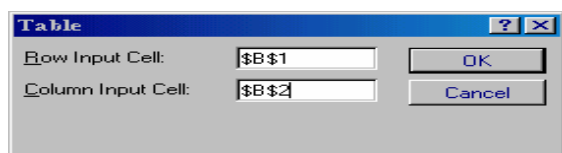
Để giải bài toán này ta làm như sau:

- Nhập số liệu và tính mẫu PMT tương tự như phần trên
- Tạo bảng tính nhanh:

	A	B	C	D	E	F	G	H	I
1	Tiền vay	125000	7.50%	30	Tiền trả hàng tháng	(\$874.02)			
2	Lãi suất				Bảng tính nhanh các mức trả hàng tháng				
3	Số năm				(với khoản tiền vay và các mức lãi suất khác nhau)				
4					Mức trả hàng tháng				
5					Mức trả hàng tháng				
6					Mức trả hàng tháng				
7					Mức trả hàng tháng				
8					Mức trả hàng tháng				
9					Mức trả hàng tháng				
10					Mức trả hàng tháng				
11					Mức trả hàng tháng				
12					Mức trả hàng tháng				
13					Mức trả hàng tháng				
14					Mức trả hàng tháng				
15					Mức trả hàng tháng				
16					Mức trả hàng tháng				
17					Mức trả hàng tháng				
18					Mức trả hàng tháng				

(Tạo miền dữ liệu cho biến thứ nhất: A10:A18; tạo miền dữ liệu cho biến thứ 2: B2:B7; Trong ô phía trên biến thứ nhất nhập vào 1 địa chỉ chứa kết quả của bước phân tích ban đầu: Tại ô A9 nhập = B5).

- Chọn vùng chứa tất cả các biến giá trị để kiểm tra và công thức địa chỉ: A9:B18
- Vào lệnh Data/Table:



- + Row Input Cell: B1 là chỉ ô ở vùng phân tích lúc đầu tương ứng với các giá trị thuộc về dòng đầu của bảng tính nhanh.
- + Column Input Cell: B2 là chỉ ra các ô vùng phân tích lúc đầu tương ứng với các giá trị thuộc về cột đầu của bảng.
- Chọn OK ta có kết quả cần tính:

The screenshot shows an Excel spreadsheet titled "Excel - TnhanhDA". The active sheet is "Sheet1". The data is organized as follows:

	A	B	C	D	E	F	G	H	I
1	Tiền vay	125000							
2	Lãi suất	7.50 %							
3	Số năm	30							
4									
5	Tiền trả hàng tháng	(=\$B74-02)							
6	Bảng tính nhanh các mức trả hàng tháng								
7	(với khoản tiền vay và các mức lãi suất khác nhau)								
8			Mức trả hàng tháng						
9	(=\$B74-02)	\$115,000.00	120000	125000	130000	135000	140000		
10	6.50 %	-726.878227	-758.4816	-790.085	-821.688	-853.29	-884.895		
11	6.75 %	-745.887811	-778.3177	-810.7476	-843.178	-875.61	-908.037		
12	7 %	-765.097869	-798.563	-831.6281	-864.893	-898.16	-931.423		
13	7.25 %	-784.502722	-818.6115	-852.7204	-886.829	-920.94	-955.047		
14	7.50 %	-804.096685	-839.0574	-874.0181	-908.979	-943.94	-978.9		
15	7.75 %	-823.874082	-859.6947	-895.5153	-931.336	-967.16	-1002.98		
16	8 %	-843.82926	-880.5175	-917.2057	-953.894	-990.58	-1027.27		
17	8.25 %	-863.956595	-901.5199	-939.0833	-976.647	-1014.2	-1051.77		
18	8.50 %	-884.250506	-922.6962	-961.1419	-999.588	-1038	-1076.48		

Chú ý:

- Các biến giá trị ta sử dụng ở cột đầu tiên của bảng hay ở dòng đầu tiên của bảng tính không nhất thiết phải tăng (giảm theo đúng số nguyên, tuy nhiên nên chọn các giá trị đó để dễ kiểm tra).
- Có thể đổi vị trí cho nhau giữa 2 biến trong bảng tính nhanh.
- Nếu thay đổi thời hạn vay (B3) thì toàn bộ bảng tính của máy sẽ tính toán lại.
- Nếu công thức tính PMT ở ô B3 thay đổi thì toàn bộ bảng tính cũng thay đổi.
- Thay đổi các giá trị của các biến thì bảng tính sẽ thay đổi giá trị.

3.3 Dùng lệnh Tools/Solve để lập kế hoạch sản xuất, phân tích sản xuất

Giả sử DN A có bảng trình bày ngân sách trong 5 năm cho sản phẩm mới như sau:

	A	B	C	D	E	F
1	Năm	1	2	3	4	5
2	Doanh số sản phẩm	50000	150000	200000	300000	500000
3	Chi phí hoạt động					
4	Quảng cáo	92392	48098	24524	24524	24524
5	Quản lý sản xuất	52699	58169	61786	66272	72489
6	Chi phí Marketing	70720	0.00B+00	0.00B+00	0.00B+00	0
7	Chi phí bảo dưỡng	0	2605	2605	2605	2605
8						
9	Tổng chi phí hoạt động	?	?	?	?	?
10						
11	EBITDA (Doanh số - tổng chi phí hoạt động)	?	?	?	?	?
12	Khấu hao nhanh	?	?	?	?	?
13	Thu nhập trước thuế	?	?	?	?	?
14	Thuế < = 36%	?	?	?	?	?
15	Thu nhập thuần	?	?	?	?	?
16	Khấu hao nhanh	?	?	?	?	?
17	Vốn đầu tư	260528	0	0	0	0
18	Dòng tiền thuần	?	?	?	?	?
19						
20	Dòng tiền thuần tích lũy	?	?	?	?	?

- Hãy tính toán các số liệu bỏ trống (“?”) trong bảng tính trên.
- Hãy xây dựng kế hoạch sản xuất sản phẩm mới sao cho đúng 4 năm là thu hồi vốn với điều kiện vốn đầu tư để sản xuất sản phẩm mới thấp nhất là \$250.000.

Để giải bài toán ta làm như sau:

- Nhập số liệu vào bảng tính Excel.
- Lập các công thức tính các chỉ tiêu kinh tế theo yêu cầu của bài toán đang đặt ra.

	B	C	D	E	F	G
1	1	2	3	4	5	
2	50000	150000	200000	300000	500000	
3						
4	92392	48098	24524	24524	24524	
5	52699	58169	61786	66272	72489	
6	70720	0.00B+00	0.00B+00	0.00B+00	0	
7	0	2605	2605	2605	2605	
8						
9	SUM(B4:B7)	SUM(C4:C7)	SUM(D4:D7)	SUM(E4:E7)	SUM(F4:F7)	
10	B2-B9	C2-C9	D2-D9	B2-B9	F2-F9	
11	DDB(\$B\$17:\$G\$12)	DDB(\$B\$17:\$G\$12)	DDB(\$B\$17:\$G\$12)	DDB(\$B\$17:\$G\$12)	DDB(\$B\$17:\$G\$12)	2
12	(B11-B12)	(C11-C12)	(D11-D12)	(B11-B12)	(F11-F12)	
13	B13*0.36	C13*0.36	D13*0.36	B13*0.36	F13*0.36	
14	B13-B14	C13-C14	D13-D14	B13-B14	F13-F14	
15	DDB(\$B\$17:\$G\$12)	DDB(\$B\$17:\$G\$12)	DDB(\$B\$17:\$G\$12)	DDB(\$B\$17:\$G\$12)	DDB(\$B\$17:\$G\$12)	
16	260528	0	0	0	0	
17	(B15+B16-B17)	(C15+C16-C17)	(D15+D16-D17)	(B15+B16-B17)	(F15+F16-F17)	
18	B18	(B\$20+C\$18)	(C\$20+D\$18)	(D\$20+B\$18)	(B\$20+F\$18)	
19						
20						

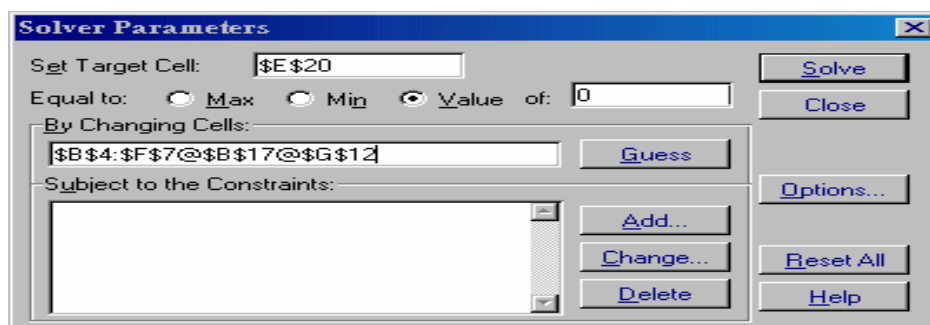
Sau khi lập công thức trên bảng tính ta được kết quả sau đây:

	A	B	C	D	E	F	G
1	Năm	1	2	3	4	5	
2	Doanh số sản phẩm	50000	150000	200000	300000	500000	
3	Chi phí hoạt động						
4	- Quảng cáo	92392	48098	24524	24524	24524	
5	- Quản lý sản xuất	52699	58169	61786	66272	72489	
6	- Chi phí Marketing	70720	0.00B+00	0.00B+00	0.00B+00	0	
7	- Chi phí bảo dưỡng	0	2605	2605	2605	2605	
8							
9	Tổng chi phí hoạt động	215811	108872	88915	93401	99618	
10							
11	BEITDA (Doanh số - tổng chi phí hoạt động)	-165811	41128	111085	206599	400382	
12	Khấu hao nhanh	\$52,105.60	\$41,684.48	\$33,347.58	\$26,678.07	\$21,342.45	2
13	Thu nhập trước thuế	-217916.6	-556.48	77737.416	179920.93	379039.55	
14	Thuế < = 36%	-78449.976	-200.3328	27985.47	64771.536	136454.24	
15	Thu nhập thuần	-139466.62	-356.1472	49751.946	115149.4	242585.31	
16	Khấu hao nhanh	\$52,105.60	\$41,684.48	\$33,347.58	\$26,678.07	\$21,342.45	
17	Vốn đầu tư	260528	0	0	0	0	
18	Dòng tiền thuần	-347889.02	41328.333	83099.53	141827.46	263927.76	
19							
20	Dòng tiền thuần tích lũy	-347889.02	-306560.69	-223461.16	-81633.697	182294.07	

Nhận xét: Ta thấy rằng dòng tiền thuần tích lũy cuối năm thứ 4 vẫn âm và sẽ dương vào năm thứ 5.

Giả sử mục tiêu đặt ra là khoảng thời gian thu hồi vốn là đúng 4 năm, doanh số bán ra không đổi, vốn đầu tư thấp nhất để sản xuất sản phẩm mới mới là 250.000. Muốn vậy ta dùng Tools/ Solver để giải quyết vấn đề này.

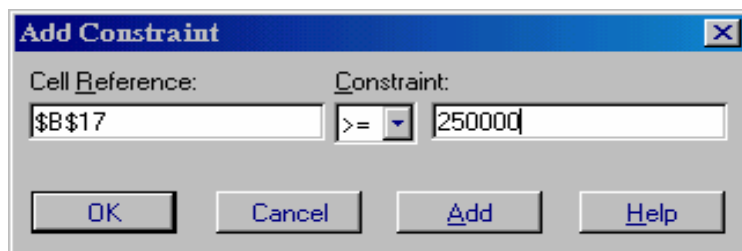
- Vào Tools/ Solver, sau đó khai báo các thông số như sau:



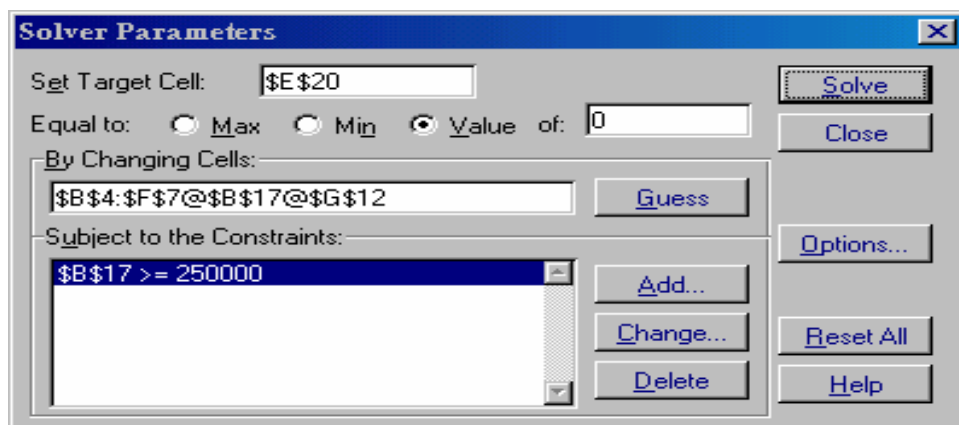
- Chọn ☒ Value bằng 0 là vì tại đó là dòng tiền tích lũy của năm thứ 4 = 0

- Nhấp lên By Changing Cells chọn B4:F6 giữ phím Ctrl và chọn tiếp B17 rồi chọn thêm ô G12 là nhằm mục đích tính toán lại phí tổn điều hành, vốn đầu tư, thay đổi hệ số khấu hao nhanh DDB bằng 2 (tức 200%).

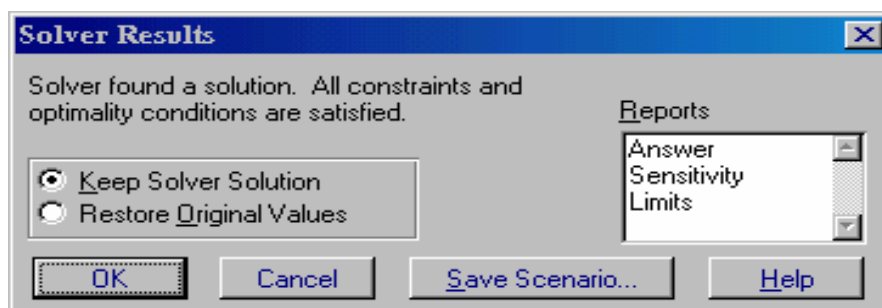
- Nhấp lên nút Add trong hộp thoại Add Constraint:



Chọn OK (Sở dĩ ta nhập B17 >= 250000 là thỏa mãn điều kiện: Vốn đầu tư thấp nhất để sản xuất kinh doanh sản phẩm mới) ta có hộp thoại:



Nhấp nút Solve để giải bài toán:



	A	B	C	D	E	F	G
1	Năm	1	2	3	4	5	
2	Doanh số sản phẩm	50000	150000	200000	300000	500000	
3	Chi phí hoạt động						
4	Quảng cáo	64054.07	40418.138	22527.442	22527.442	24524	
5	Quản lý sản xuất	43479.569	46936.341	49112.997	51691.933	72489	
6	Chi phí Marketing	54117.103	-3.32E-06	-3.32E-06	-3.32E-06	0	
7	Chi phí bảo dưỡng	-3.321E-06	2582.4724	2582.4724	2582.4724	2605	
8	Tổng chi phí hoạt động	161650.74	89936.951	74222.911	76801.847	99618	
9	BBITDA (Doanh số - tổng chi phí hoạt động)	-111650.74	60063.049	125777.09	223198.15	400382	
10	Khấu hao nhanh	\$59,523.93	\$45,351.54	\$34,553.53	\$26,326.48	\$20,058.26	2.381
11	Thu nhập trước thuế	-171174.67	14711.511	91223.558	196871.67	380323.74	
12	Thuế <= 36%	-61622.883	5296.1439	32840.481	70873.802	136916.55	
13	Thu nhập thuần	-109551.79	9415.3669	58383.077	125997.87	243407.19	
14	Khấu hao nhanh	\$59,523.93	\$45,351.54	\$34,553.53	\$26,326.48	\$20,058.26	
15	Vốn đầu tư	250000	0	0	0	0	
16	Dòng tiền thuần	-300027.86	54766.906	92936.608	152324.35	263465.45	
17							
18	Dòng tiền thuần tích lũy	-300027.86	-245260.95	-152324.34	0.0073968	263465.46	

Ta nhận thấy rằng cuối năm thứ 4 đã thu hồi vốn, các chi phí thay đổi nhưng duy nhất doanh số không thay đổi.

(Tương tự như vậy, ta có thể lập và phân tích kinh tế sản xuất để đúng 3 năm sản xuất kinh doanh sản phẩm mới sẽ thu hồi vốn).

BÀI TẬP THỰC HÀNH CHƯƠNG II

Bài 1. Cho nguồn thông tin

DT (ha)	10	20	25	30	40	45	50	55	60	65
NS (tạ/ ha)	42	45	35	42	43	46	50	47	48	49

1- Tính sản lượng theo NS và DT.

2- Tính NS bình quân chung đối với tổng toàn bộ diện tích

3- Giả sử có mô hình: $NS = a + b.DT$, hãy dùng hàm Intercept và Slope để tính các hệ số ước lượng a, b.

4- Dùng công thức thống kê để tính các hệ số ước lượng a, b.

5- Dùng lệnh Tools/ Data Analysis/ Regression để chạy mô hình hồi quy trên. Cho nhận xét về mô hình?

Bài 2. Cho thông tin của 1 doanh nghiệp như sau:

Q (tấn)	0	10	20	30	40	50	60	65	70	75
P (ngđ/kg)	15	14	13.5	13	12.6	12.3	12.1	11.8	11.5	11
AVC (ngđ/kg)	0	9	8.8	8.6	8.5	8.4	8.3	8	7.8	7.5

FC = 500 trđ

- 1- Tính VC, TC, AFC, ATC, MC theo Q và AVC
- 2- Tính Q hòa vốn
- 3- Tính TR, MR, tìm điểm Pr max
- 4- Vẽ đồ thị TC, TR trên cùng trục tọa độ
- 5- Hãy chạy mô hình hồi quy $Q = a.ATC + b.P$, cho nhận xét.

Bài 3. Thống kê số lao động trong 1 DN như sau:

a) Ngày	1/1	10/1	16/1	20/1	26/1	31/1
Số người LĐ	100	120	125	115	130	125

Tính số lao động bình quân trong tháng 1 cho DN trên.

b) Ngày	1/1	1/2	1/3	1/4	1/5	1/6	1/7	1/8	1/9	1/10	1/11	1/12	31/12
Số người LĐ	100	105	110	115	120	120	125	125	130	130	130	115	120

- 1- Tính số lao động bình quân từng tháng trong năm
- 2- Tính số lao động bình quân quý trong năm
- 3- Tính số LĐ bình quân 6 tháng đầu năm và 6 tháng cuối năm
- 4- Tính số LĐ bình quân 1 tháng trong năm

Bài 4. Cho bảng kết quả sản xuất tại 1 cơ sở (ĐVT: trđ)

Diễn giải	2003	2004	2005	2006	2007
1. Nông nghiệp					
- Trồng trọt	300	320	350	370	400
+ Lúa	180	175	200	200	250
- Chăn nuôi	120	125	150	150	150
+ Lợn	70	75	80	90	100
2. Công nghiệp TM	160	165	180	200	200
3. Khác	40	40	50	50	60

- 1- Tính kết quả SX cho ngành NN ?
- 2- Tính tổng toàn bộ kết quả sản xuất hàng năm cho cơ sở ?
- 3- Tính tốc độ phát triển liên hoàn, phát triển định gốc và phát triển bình quân về kết quả sản xuất cho cơ sở ?
- 4- Xác định cơ cấu KT cho cơ sở ?

Bài 5. Thống kê tài sản cố định của 1 DN như sau:

Loại TSCĐ	Giá trị (trđ)	Thanh lý (trđ)	Số năm SD (năm)
A	200	15	10
B	500	40	15
C	1000	100	15
D	1500	85	12
E	600	45	8

- 1- Tính tỷ lệ khấu hao bình quân hàng năm cho từng TS.
- 2- Tính mức KH (trđ) hàng năm cho từng loại TS theo các PP đường thẳng, tổng số số năm sử dụng, số dư giảm dần.
- 3- Tính tổng mức KH hàng năm theo P đường thẳng, từ đó tính thời gian sử dụng bình quân chung cho các loại TS của DN.

4- Tính tỉ lệ KH bình quân cho các loại TS của DN theo PP đường thẳng ?

Bài 6. Một trang trại pha trộn phân bón để đảm bảo bón đủ cho nhu cầu cây trồng với lượng tối thiểu là ka li = 15 đơn vị, đạm = 20 đơn vị, lân = 24 đơn vị. Phân được pha trộn từ 2 loại sau:

- Loại 1 có chứa 3 đơn vị ka li, 1 đơn vị đạm, 3 đơn vị lân. Giá của phân loại 1 là 320 ngàn đồng/ tạ
- Loại 2 có chứa 1 đơn vị ka li, 5 đơn vị đạm và 2 đơn vị lân. Giá của phân loại 2 là 380 ngàn đồng/ tạ

Hãy kết hợp trộn 2 loại phân này nhằm đủ chất dinh dưỡng cho cây trồng nhưng giá thành nhỏ nhất.

Bài 7. Dùng hàm tài chính để lựa chọn các thông tin sau:

Khoản mục	Dự án A	Dự án B	Dự án C
- Nguyên giá (trđ)	500	450	480
- Thanh lí (trđ)	25	20	24
- Số năm SD (năm)	20	18	19
- Số năm/ lần đại tu (năm)	9	10	8
- Tiền đại tu/ lần (trđ)	80	40	50
- CP thường xuyên/ năm (trđ)	150	120	140
- Thu nhập hàng năm (trđ)	200	180	195

- 1- Dùng hàm PV để lựa chọn dự án A, B, C ?
 - 2- Dùng hàm FV để lựa chọn dự án A, B, C ?
 - 3- Dùng hàm NPV và IRR để chọn dự án A, B, C ?
- Cho biết suất chiết khấu là 9%/ năm.

Bài 8. Tính khấu hao tài sản cố định

Một xí nghiệp mua một dây chuyền sản xuất hết 300 triệu đồng. Thời gian sử dụng của dây chuyền theo thiết kế là 12 năm. Giá trị thanh lý của dây chuyền dự đoán còn 10 triệu đồng. Yêu cầu: Tính khấu hao của dây chuyền theo các phương pháp khác nhau cho năm thứ 1, 3, 7, 8 và năm cuối cùng (hàm SLN, DB, DDB, SYD).

Bài 9. Có tài liệu về các dự án đầu tư như sau:

Chỉ tiêu	Dự án A				Dự án B			
	Năm 0	Năm 1	Năm 2	Năm 3	Năm 0	Năm 1	Năm 2	Năm 3
Sản lượng SP (tấn)		5.4	7.1	9.2		8.3	7.1	6.4
Giá (tr.đ/tấn)		5.5	6.4	7.2		5.3	6.2	7.6
Chi phí (tr.đ)	30.0	9.3	15.0	18.0	30.0	17.0	16.0	8.2
Thuê (tr.đ)	5.2	8.6	9.8	9.7	5.2	8.4	10.0	9.5

1. Tính thu nhập của các dự án
2. Tính PV của các dự án theo các năm (biết rằng lãi suất dự kiến là 10%/năm)
3. Trong 2 dự án trên dự án nào có hiệu quả hơn.

Bài 10. Có số liệu về 1 dự án như sau (ĐVT: tr.đ)

Vốn xây dựng cơ bản	100
Lãi thu năm 1	10
Lãi thu năm 2	20
Lãi thu năm 3	30
Lãi thu năm 4	40
Lãi thu năm 5	50
Lãi thu năm 6	55
Lãi thu năm 7	60
Lãi thu năm 8	65
Lãi thu năm 9	70

1. Tính NPV của dự án, lãi suất ngân hàng bình quân là 9%/năm
2. Hãy tính IRR (giả sử dự án kết thúc ở năm thứ 9). Nếu như dự án kết thúc sớm hơn, hãy dự đoán khoảng thời gian hoàn vốn.

Bài 11. Tính giá trị tương lai của dòng tiền

Bà An có 100 triệu đồng gửi tiết kiệm với lãi suất 0,7%/tháng. Hỏi sau 3 năm bà ấy sẽ nhận về số tiền là bao nhiêu? Nếu như lãi suất thay đổi thành 2,5%/quý (3 tháng) thì số tiền bà An sẽ nhận được là bao nhiêu với thời gian trên (3 năm).

Bài 12. Tính giá trị hiện tại của dòng tiền

Một dự án khai thác và nuôi trồng thủy hải sản bắt đầu vào tháng 1/2001. Dự kiến kết thúc tháng 1/2007 khi đó giá trị sản phẩm thu hoạch sẽ là 1 tỷ đồng. Hàng năm (từ 2001 – 2007) dự án thu thêm được 100 triệu đồng tiền bán sản phẩm. Hãy tính giá trị hiện tại của toàn bộ sản phẩm mà dự án thu được. Biết rằng lãi suất ngân hàng bình quân là 10%/năm. Nếu như sản phẩm bây giờ thu hàng tháng, mỗi tháng 10 triệu đồng, lãi suất 0,8%/tháng thì tổng giá trị hiện tại của sản phẩm dự án thu được bây giờ sẽ là bao nhiêu?

Bài 13. Cho trước 1 số thông tin, hãy hoàn thiện bảng sau:

Phương án sản xuất 1 loại bánh trung thu

Diễn giải	ĐVT	Định mức	Số lượng	Đơn giá	Thành tiền	Tỉ lệ
Nguyên liệu chính					?	?
- Đường	%	15	?	60	?	?
- Thịt	%	12	?	50	?	?
- Đậu	%	10	?	30	?	?
- Trứng	%	10	?	60	?	?
- Bơ	%	5	?	80	?	?
- Bột mì	%	?	?	20	?	?
Vật liệu phụ					?	?
- Dầu mỡ	%	5	?	40	?	?
- Gia vị	%	3	?	80	?	?
Chi phí sản xuất					?	?
- Chất đốt	%	8	?	10	?	?
- Bao bì	%	2	?	20	?	?
- Công lao động	%				?	12%
- Chi phí chuyên chở	%				?	3%
- Chi phí bảo quản	%				?	2%
Chi phí khác					?	?
- Thuế kinh doanh	%				?	3%
- Hoa hồng đại lí	%				?	2%
- Quảng cáo tiếp thị	%				?	1%
- Sản lượng sản xuất	kg	2000		?	?	

Bài 14. Bảng kiến thức kinh tế và tin học ứng dụng, hãy hoàn thiện bảng thanh toán công tác phí sau đây:

STT	Họ và tên	Chức vụ	Phụ cấp	Cụ ly	Phân loại cấp Km thực tế			Tổng tiền công tác phí
			(đ/km)	(km)	Đến	Đến	Trên	(đ)

					50	120	120	
1	Lê Minh Sơn	PP	?	150	?	?	?	?
2	Ngô Văn Hùng	NV	?	40	?			?
3	Vũ Minh Ngọc	GD	?	100	?	?		?
4	Trần Văn Nam	TP	?	160	?	?	?	?
5	La Thu Thủy	PP	?	200	?	?	?	?
6	Thân Văn An	NV	?	45	?			?
7	Hà Văn Kiệt	NV	?	120	?	?		?
8	Cần Đại Dương	TP	?	220	?	?	?	?
Tổng cộng Bình quân Cao nhất Thấp nhất				?	?	?	?	?
				?	?	?	?	?
				?	?	?	?	?
				?	?			?

Cho biết phụ cấp chức vụ và phụ cấp đường xá như sau:

Chức vụ GD TP PP NV; Cự ly (km) Đến 50 Đến 120 Trên
120
Phụ cấp chức vụ (đ) 400 350 300 250; Phụ cấp đường xá (%) 100% 120%
150%

Bài 15. Bằng kiến thức kinh tế và tin học ứng dụng đã được học, hãy hoàn thành bảng phân bổ vật tư các công trình sau đây:

Số TT	Tên vật tư	Đơn giá (đồng)	Đơn vị tính	Công trình A		Công trình B		Công trình C		Cộng thành tiền (đ)	Tỉ lệ (%)
				Số lượng	Giá trị (đ)	Số lượng	Giá trị (đ)	Số lượng	Giá trị (đ)		
1	Xi măng	49.000	bao	50	?	30	?	40	?	?	?
2	Sắt L50	3.000	kg	120,5	?	80	?	100	?	?	?
3	Sắt tròn	3.200	kg	150	?	90	?	60	?	?	?
4	Sơn HN	14.500	kg	25	?	18,5	?	13,5	?	?	?
5	Sơn ch.thấm	15.000	kg	12	?	10	?	11,5	?	?	?
6	Gỗ 5x10	12.000	mét	12,5	?	14	?	11	?	?	?
7	Cát	1.400	m ³	12	?	15	?	9	?	?	?
8	Đá 1x2	18.000	m ³	10	?	11	?	14	?	?	?
Tổng cộng					?		?		?	?	
Tỉ lệ (%)					?		?		?		

Bài 16. Ứng dụng tin học để hoàn thành bảng kết quả kết quả học tập của 1 tổ sau đây:

Họ tên SV	Khu Vực	Giới tính	Điểm m T.kê	Điểm m Toán	Điểm K.toán	Điểm Ngngữ	Điểm Tin UD	BQ chung	Điểm ưu tiên	Kết quả	Vị thứ	Xếp loại	Điểm m thấp	Môn thấp
Lê Văn Thanh	B	Nam	4	5	7	9	6	?	?	?	?	?	?	?
Nguyễn Ngọc	A	Nữ	5	6	7	7	8	?	?	?	?	?	?	?
Vũ Đức Minh	C	Nam	8	8	9	7	9	?	?	?	?	?	?	?
Phạm Văn Đạt	C	Nam	9	9	8	6	9	?	?	?	?	?	?	?
Bùi Ngọc Mai	B	Nữ	10	8	9	7	8	?	?	?	?	?	?	?
Phí Tiến Mạnh	A	Nam	6	7	8	8	10	?	?	?	?	?	?	?
Lê Minh Sơn	A	Nam	7	7	8	7	9	?	?	?	?	?	?	?
Ngô Văn Hùng	B	Nam	8	8	9	8	9	?	?	?	?	?	?	?
Vũ Minh Ngọc	A	Nữ	9	8	8	8	7	?	?	?	?	?	?	?
Trần Văn Nam	C	Nam	7	10	8	7	9	?	?	?	?	?	?	?
La Thu Thủy	C	Nữ	5	6	7	9	8	?	?	?	?	?	?	?
Thân Văn An	B	Nam	3	8	7	8	7	?	?	?	?	?	?	?
Hà Văn Kiệt	A	Nam	7	5	6	7	6	?	?	?	?	?	?	?
Cần Đại Dương	A	Nam	8	6	7	8	8	?	?	?	?	?	?	?
Trung bình			?	?	?	?	?	?	?	?				
Cao nhất			?	?	?	?	?	?	?	?				
Thấp nhất			?	?	?	?	?	?	?	?				

* Điểm BQ chung được tính theo số trình của môn học như sau:

Môn học: Thống kê Toán Kế toán Ngoại ngữ Tin UD

Số trình học: 4 4 4 3 3

* Điểm ưu tiên khu vực như sau:

Khu vực: A B C

Điểm ưu tiên: 0 0,25 0,4

* Cách xếp loại theo kết quả như sau:

Kết quả Dưới 5 5 – 6 Trên 6, dưới 7 Từ 7 đến dưới 8 Từ 8 đến dưới 9 Từ 9 trở lên

Loại yếu TB TB khá Khá Giỏi Xuất sắc

* Kết quả = BQ chung + Ưu tiên

* Vị thứ: Xếp hạng theo kết quả, người có kết quả cao nhất là số 1

* Điểm ưu tiên giới tính như sau:

Giới tính : Nam Nữ

Điểm ưu tiên: 0 0,25

Bài 17. Công ty A sản xuất gạch chịu nhiệt ước tính chi phí biến đổi để sản xuất 1 tấn gạch chịu nhiệt là 2,7 triệu đồng, giá bán bình quân trên thị trường là 3,4 triệu đồng, được biết định phí hàng tháng của công ty là 650 triệu đồng. Hãy thiết lập bảng tính để phân tích điểm hoà vốn bằng công cụ Goal Seek của phần mềm EXCEL

Bài 18. Một công ty du lịch chuyên kinh doanh dịch vụ khách sạn đã lập báo cáo doanh số tháng 1/2004 như sau:

Báo cáo doanh số tháng 1/2004

STT	Tên khách	Mã quốc tịch	Quốc tịch	Loại phòng	Đơn giá	Ngày đến	Ngày đi	Thời gian ở	Thành tiền	Giảm	Phải trả
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1	Anh	5		1		1/1/03	30/1/03				
2	Hùng	2		4		2/1/03	1/2/03				
3	Minh	3		1		3/1/03	17/1/03				
4	Ngọc	1		1		4/1/03	21/1/03				
5	Đức	3		4		5/1/03	25/1/03				
6	Hồng	1		2		6/1/03	3/2/03				
7	Hải	2		1		7/1/03	15/1/03				
8	Mạnh	5		3		8/1/03	13/1/03				
9	Tiến	5		3		9/1/03	3/2/03				
10	Phú	2		3		10/1/03	27/1/03				

Cho biết quốc tịch, đơn giá loại phòng, tỷ lệ giảm như sau:

Mã quốc tịch	1	2	3	4	5
Quốc tịch	úc	Trung quốc	Nga	Pháp	Mỹ
Loại phòng	1	2	3	4	5
Đơn giá (\$/ngày)	300	250	200	150	100
Thời gian ở (ngày)	1 → 4	5 → 9	10 → 14	15 → 19	20 trở lên
Tỷ lệ giảm (%)	0	2	4	6	8

Biết rằng thời gian ở tính từ ngày đến tới ngày trả phòng.

- Dùng các hàm VLOOKUP, SUM, IF và các phép tính phù hợp cho từng cột (4, 6, 9, 10, 11, 12) để xác định kết quả cho bảng báo cáo đã nêu trên.
- Dùng hàm COUNTIF, SUM, SUMIF và các phép tính phù hợp để lập bảng tổng hợp số tiền thuê và số lượt người theo quốc tịch người thuê, loại phòng, thời gian thuê dưới dạng bảng sau:

Tổng hợp số liệu											
Theo quốc tịch				Theo loại phòng				Theo thời gian ở			
Quốc tịch	Số lượt người	Tiền	Tỷ lệ	Loại phòng	Số lượt người	Tiền	Tỷ lệ	Thời gian	Số lượt người	Tiền	Tỷ lệ
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)

- Bạn có nhận xét gì khi giải quyết các yêu cầu trên ?

Bài 19. Giả sử doanh nghiệp A muốn tận dụng thời gian nhàn rỗi của máy để sản xuất thêm 2 sản phẩm B, C. Mỗi sản phẩm cần sử dụng 2 máy. Hiện tại doanh nghiệp có 1 máy có thể sử dụng 12 giờ, máy kia có thể sử dụng 8 giờ. Mỗi đơn vị sản phẩm B yêu cầu thời gian 2 giờ cho mỗi máy. Mỗi đơn vị sản phẩm C yêu cầu 3 giờ đối với máy 1 và 1 giờ đối với máy 2. Lợi nhuận/ 1 đơn vị sản phẩm P là 6 ngàn đồng và sản phẩm C là 7 ngàn đồng. Doanh nghiệp có thể bán hết các sản phẩm hàng hoá được sản xuất. Hãy bố trí sản xuất cho doanh nghiệp nhằm mang lại lợi nhuận cực đại.

Bài 20. Có số liệu về tình hình đầu tư, chi phí vận hành và doanh thu của việc sử dụng 2 loại thiết bị cùng công dụng như sau:

Diễn giải	ĐVT	Thiết bị A	Thiết bị B
Vốn đầu tư ban đầu	Trđ	7	12
CP vận hành/ năm	Trđ	2	3
Doanh thu/ năm	Trđ	5	7
Giá trị còn lại	Trđ	1	0
Năm hoạt động	Năm	3	4

1- Hãy lựa chọn phương án thiết bị theo chỉ tiêu thu nhập thuần, biết rằng tỷ suất lợi nhuận định mức là 10%.

2- Hãy chọn phương án thiết bị theo tỷ suất nội hoàn vốn

Bài 21. Hoàn thành bảng báo cáo thu nhập theo tỉ lệ chung sau đây (ĐVT: trđ):

Diễn giải	2002		2003		2004		So sánh (%)	
	Số tiền	Tỉ lệ (%)	Số tiền	Tỉ lệ (%)	Số tiền	Tỉ lệ (%)	2003/2002	2004/2003
- Doanh thu	9.734	?	8.028	?	7.841	?	?	?
- Giá vốn hàng bán	6.085	?	4.843	?	4.648	?	?	?
- Lãi gộp	?	?	?	?	?	?	?	?
- CP bán hàng	1.030	?	891	?	868	?	?	?
- CP quản lí	602	?	527	?	500	?	?	?
- Lợi nhuận từ HĐKD	?	?	?	?	?	?	?	?
- Thu nhập lãi vay	90	?	84	?	86	?	?	?
- CP lãi vay	345	?	314	?	342	?	?	?
- Lợi nhuận trước thuế	?	?	?	?	?	?	?	?
- Thuế thu nhập (34%)	?	?	?	?	?	?	?	?
- Lợi nhuận ròng	?	?	?	?	?	?	?	?

Sau khi tính toán xong, bạn hãy phân tích khái quát báo cáo thu nhập ?

Bài 22. Cho các thông tin sau đây liên quan đến dự án đầu tư đang được công ty A xem xét:

- Kinh phí đầu tư ban đầu: 80.000 trđ - Đầu tư vốn lưu động bổ sung: 10.000 trđ

- Thu nhập thuần hàng năm: 25.000 trđ - Giá trị thanh lí TSCĐ: 2.500 trđ

Giả định thời gian của dự án đầu tư là 4 năm và tỷ suất chiết khấu thị trường là $r = 8\%$, công ty này có nên đầu tư vào dự án không ?

Bài 23. Công ty in Hồng Hà đang phải đối mặt với một vấn đề là phải thay thế những máy in đã cũ. Có hai phương án lựa chọn đều đáp ứng yêu cầu của công ty như nhau. Phương án A chi phí là 50.000\$ và đòi hỏi chi phí vận hành hàng năm là 20.000\$. Phương án B chi phí 75.000\$, nhưng chi phí vận hành hàng năm chỉ hết 15.000\$. Cả hai phương án đều có tuổi thọ là 10 năm. Cả hai máy đều khấu hao theo PP tổng số số năm sử dụng. Tuy nhiên, lựa chọn A có giá trị thanh lí là 5.000\$ và B là 10.000\$. Hãy đưa ra lời khuyên cho công ty in Hồng Hà về chiến lược đầu tư, giả định chi phí cơ hội của vốn là 8%; 10% ?

Bài 24. Hoàn thành bảng lương tháng 1/2005 của công ty X:

STT	Họ và tên	Mã bậc lương	Lương chính	Mã chức vụ	Chức vụ	Phụ cấp	Công tạm ứng	Tiền tạm ứng	Công thực làm	Lương	Thưởng	BHXH	BHYT, công đoàn	Thuế thu nhập	Còn lại
1	Trần Trọng Hiếu	1	?	KT	?	?	13	?	28	?	?	?	?	?	?
2	Lê Khắc Chiến	2	?	NV	?	?	11	?	26	?	?	?	?	?	?
3	Phạm Thanh	5	?	PGĐ	?	?	11	?	25	?	?	?	?	?	?
4	Hoàng Hồng	2	?	KT	?	?	10	?	29	?	?	?	?	?	?
5	Lê Minh	2	?	KT	?	?	12	?	25	?	?	?	?	?	?
6	Bùi Minh Vũ	4	?	TP	?	?	10	?	24	?	?	?	?	?	?
7	Trương Ngọc	3	?	PP	?	?	11	?	25	?	?	?	?	?	?
8	Lã Thanh Thu	6	?	GĐ	?	?	9	?	25	?	?	?	?	?	?
9	Lương Đức Ty	1	?	NV	?	?	13	?	24	?	?	?	?	?	?
10	Hà Ngọc Huy	2	?	KT	?	?	10	?	24	?	?	?	?	?	?
Tổng cộng			?			?	?	?	?	?	?	?	?	?	?
Lương cao nhất			?			?	?	?	?	?	?	?	?	?	?
Lương thấp nhất			?			?	?	?	?	?	?	?	?	?	?
Lương bình quân			?			?	?	?	?	?	?	?	?	?	?

Cho biết bảng danh mục lương của công ty X như sau:

Mã chức vụ	GĐ	KT	NV	PGĐ	PP	TP
Chức vụ	Giám đốc	Kỹ thuật	Nhân viên	Phó giám đốc	Phó phòng	Trưởng phòng
Phụ cấp	400	100	-	300	150	200
Mã bậc lương	1	2	3	4	5	6
Bậc lương	1000	1100	1200	1300	1400	1500

Quỹ lương của công ty: 50 triệu đồng/ tháng; hệ số lương là 3.000 đồng; Số ngày làm việc chuẩn: 26 ngày/ tháng

Với quy định cách thức tính lương của công ty như sau:

- Ngày công thực tính = Ngày công thực làm, nếu ngày công thực làm $\leq$ số ngày làm việc chuẩn trong tháng. Mỗi ngày công thực làm vượt quá số ngày làm việc chuẩn sẽ được tính gấp đôi.
- Lương chính = Bậc lương * Hệ số lương * (Ngày công thực tính/ Số ngày làm việc chuẩn)
- Phụ cấp = (Mức phụ cấp * Hệ số lương) * (Ngày công thực tính/ Số ngày làm việc chuẩn), mức phụ cấp căn cứ vào chức vụ nhân viên đó đảm nhận.
- Lương = Lương chính + Phụ cấp
- Tổng thưởng = Quỹ lương – Tổng lương toàn nhân viên. Doanh nghiệp tự phân bổ quỹ lương dựa vào tình hình kinh doanh trong tháng.
- Thưởng mỗi người = (Tổng thưởng/ Tổng số ngày công thực tính) * Ngày công thực tính của mỗi người. Riêng người cuối danh sách = Tổng thưởng – Thưởng của tất cả nhân viên còn lại.
- Tổng lương mỗi người = Lương + Thưởng
- Hàng tháng nhân viên được ứng lương vào thời điểm giữa tháng. Số tiền lương ứng căn cứ vào số ngày công đã làm (tạm ứng), mức lương chính và phụ cấp của từng nhân viên đến thời điểm đó.
- Bảo hiểm XH = 5% lương không kể thưởng do nhân viên chịu. Ngoài ra, DN chịu thêm 15%. Hàng tháng DN phải nộp 17% cho cơ quan BHXH, 3% DN giữ lại để chi trực tiếp cho công nhân viên về các khoản trợ cấp, ốm đau, thai sản, tai nạn lao động.

10. Bảo hiểm y tế = 1% lương không kể thưởng do nhân viên chịu. Ngoài ra, doanh nghiệp chịu thêm 2%.

11. Công đoàn phí = 1% lương không kể thưởng do nhân viên chịu.

12. Thuế thu nhập dựa trên tổng lương (kể cả thưởng), được tính theo phương pháp lũy tiến như sau:

- Dưới 5 triệu đồng thì mức thuế thu nhập = 0%
- Từ 5 đến 10 triệu đồng thì mức thuế thu nhập là 5%
- Từ 10 đến 20 triệu đồng thì mức thuế TN = 10%
- Trên 20 triệu đồng thì mức thuế thu nhập là 15%

13. Còn lại = Tổng lương – Tạm ứng – BHXH – BHYT – Công đoàn phí – Thuế thu nhập.

Bài 25. Có tài liệu sau

Năng suất lúa (tạ/ha)	Đạm (kg/ha)	Lân (kg/ha)	Phân chuồng (tấn/ha)
25.5	17.4	53.0	10.8
31.2	23.2	54.0	9.4
25.9	22.2	64.0	7.2
38.9	20.5	46.0	8.5
18.4	21.9	66.0	9.4
26.7	22.2	58.0	9.9
26.4	21.0	66.2	8.0
25.9	23.2	77.2	9.1
32.0	24.8	54.3	8.7
39.7	27.3	56.8	9.2

- Tính các chỉ tiêu thống kê mô tả về năng suất, mức bón đạm, lân, phân chuồng.
- Phân tích hồi quy và tương quan giữa năng suất với mức bón đạm, lân, phân chuồng. Theo anh (chị) có tăng mức bón đạm, lân, phân chuồng cho lúa nữa không ? Vì sao.

Chương III. ỨNG DỤNG SPSS PHÂN TÍCH DỮ LIỆU TRONG KINH TẾ VÀ QUẢN TRỊ KINH DOANH

1. Khảo sát phân phối của một biến và mối quan hệ giữa các biến

1.1. Biến định tính

- Khi biến được nghiên cứu mà giá trị số liệu là danh nghĩa, phân loại, hạng (non-numeric categories) thì SPSS xếp loại biến này vào kiểu biến định tính Nominal và Ordinal.
- Đối với số liệu danh nghĩa (nominal data) thì tần suất (Frequency), bảng chéo (Crosstabs), biểu đồ thanh (Bar charts) và biểu đồ hình bánh (Pie charts) là công cụ được sử dụng phổ biến trong SPSS.
- Đối với số liệu thứ bậc (Ordinal data) thì tần suất (Frequency), bảng chéo (Crosstabs), thống kê mô tả (Descriptive statistics), biểu đồ thanh (Bar charts), biểu đồ hình bánh (Pie charts), biểu đồ thân lá (Stem-leaf plots) là các công cụ được sử dụng phổ biến.
- Tạo bảng tần suất: Mở tệp word95.sav có sẵn trong phần mềm SPSS, chọn Analyze/Descriptive Statistics/Frequencies. Chọn biến region và chuyển sang hộp Variable(s)/OK

Statistics

Region or economic group

N	Valid	109
	Missing	0

Region or economic group

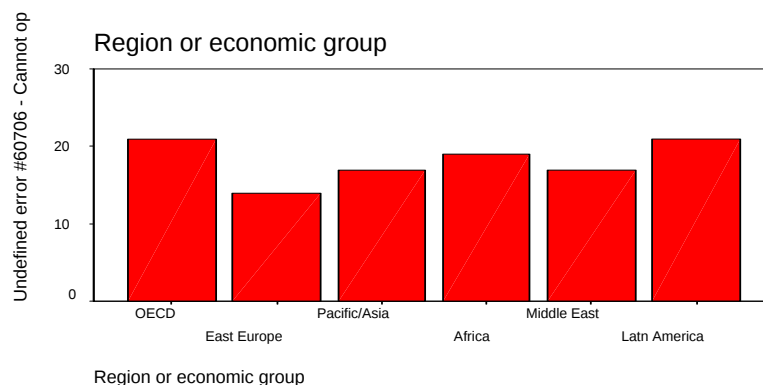
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	OECD	21	19.3	19.3	19.3
	East Europe	14	12.8	12.8	32.1
	Pacific/Asia	17	15.6	15.6	47.7
	Africa	19	17.4	17.4	65.1
	Middle East	17	15.6	15.6	80.7
	Latn America	21	19.3	19.3	100.0
	Total	109	100.0	100.0	

Bạn được bảng tần suất với các kết quả tần suất (Frequency), phần trăm (Percent), phần trăm hợp lệ (Valid Percent), phần trăm tích lũy (Cumulative Percent). Đối với dữ liệu định tính thì bảng tần suất là rất quan trọng để tóm lược.

- Tạo biểu đồ thanh để biểu diễn phân phối

Chọn Analyze/Descriptive Statistics/Frequencies (như trên)/ nhấp nút Chart/ chọn Charts/OK.

Bạn có thể tạo các bảng tần suất và đồ thị một lần cho nhiều biến bằng cách chuyển nhiều biến sang hộp Variables.

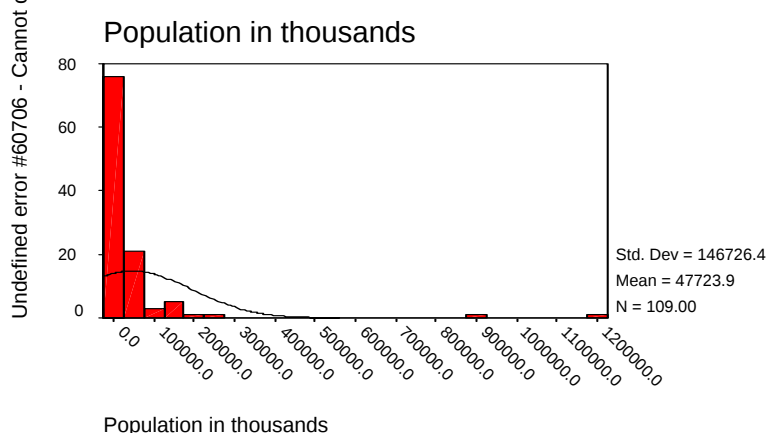


1.2. Biến định lượng

- Khi biến được nghiên cứu mà giá trị số liệu được ghi nhận bằng cách đếm hoặc đo đạc thì SPSS xếp loại biến này vào kiểu biến định lượng Scale. Một biến định lượng có thể là rời rạc (Discrete) như số sinh viên nhận được điểm khá giỏi trong lớp, số thành viên trong gia đình, hoặc liên tục (continuous) như tiền lương, chi phí, v.v...

- Đối với số liệu khoảng cách và tỷ lệ (Scale data) thì thống kê mô tả (Descriptive statistics), lược đồ tần suất (Histograms), biểu đồ hình hộp (Box plots) và biểu đồ phân tán (Scatter plots) là những công cụ đặc trưng được sử dụng phổ biến khi phân tích biến định lượng.

+ Tạo lược đồ tần suất (Histogram): Mở tệp tin word95.sav có sẵn trong phần mềm SPSS, chọn Analyze/ Descriptive Statistics/ Frequencies, chuyển biến populatn trong hộp Frequencies vào hộp Variables, chọn ô Display Frequency Tables (mặc định), nhấp nút Charts. Trong hộp thoại Frequencies: Charts bạn chọn Histograms và nhấp chọn ô With normal curve/OK.



+ Tạo biểu đồ thân lá: Chọn Analyze/ Descriptive Statistics/Explore, chuyển biến populatn vào hộp Dependent List/OK

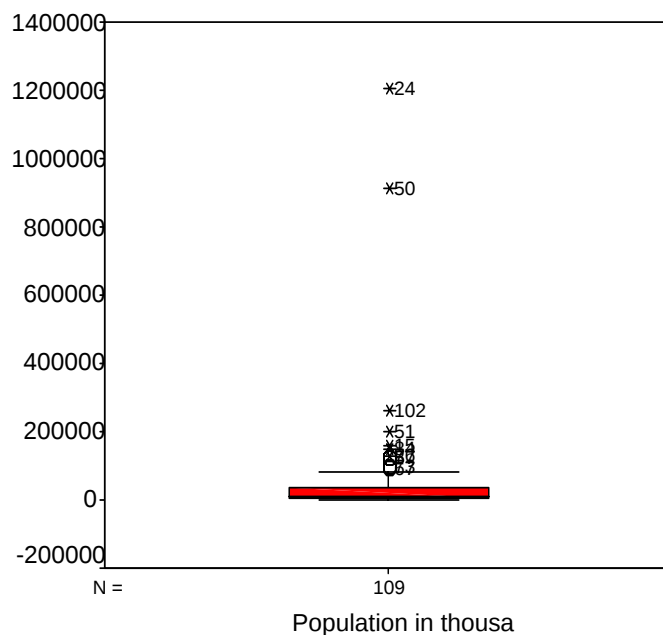
Population in thousands Stem-and-Leaf Plot

Frequency	Stem & Leaf
27.00	0 . 000011111222223333333334444
22.00	0 . 5555555556667777888889
14.00	1 . 000000000001344
6.00	1 . 578999
7.00	2 . 0002333
4.00	2 . 8899
1.00	3 . 3
3.00	3 . 589

```

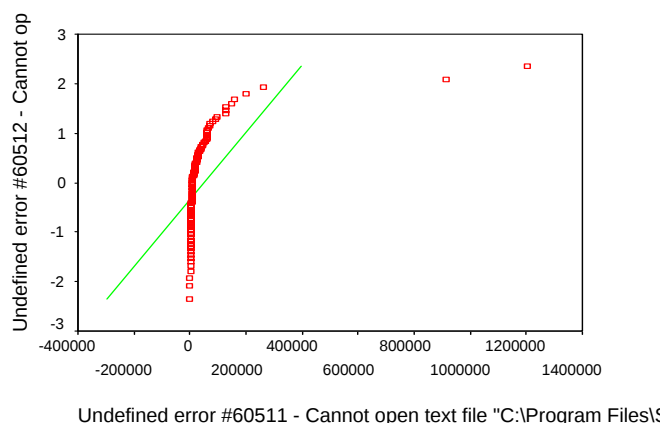
1.00      4 . 3
1.00      4 . 5
1.00      5 . 1
5.00      5 . 58889
2.00      6 . 02
2.00      6 . 59
1.00      7 . 3
.00       7 .
1.00      8 . 1
11.00 Extremes      (>=91800)
    
```

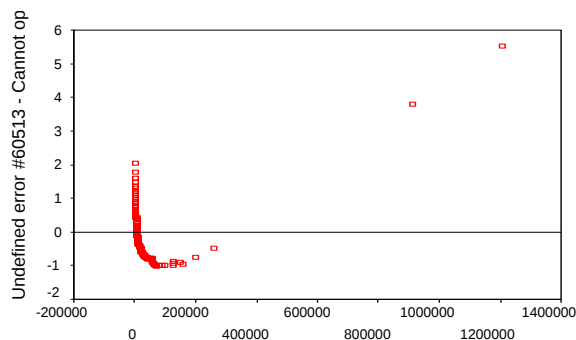
Stem width: 10000
Each leaf: 1 case(s)



Khi so sánh những biểu đồ hộp, bạn nên xem xét hình dạng, những giá trị tiêu biểu, độ phân tán và những đường nét đối với tiêu thức phân tổ. Chúng ta cũng có thể dùng biểu đồ than lá để so sánh. Tuy nhiên, biểu đồ hộp là dạng biểu diễn thuận lợi nhất.

- Tạo biểu đồ xác suất chuẩn (Q-Q plot) nhằm xem phân phối của dữ liệu có chuẩn không: Chọn Analyze/ Descriptive Statistics/Explore, chuyển biến populatn vào hộp Dependent List, chọn Plots, chọn Histogram trong phần Explore: Plots, chọn nút Normality plot with test/ Continue/OK





Undefined error #60511 - Cannot open text file "C:\Program Files\SPS:

1.3. Mối liên hệ giữa biến định tính và định lượng

a) Mối quan hệ giữa các biến định tính

* Mối quan hệ giữa 2 biến định tính

- Dùng bảng chéo (crosstab) (bảng phân tổ kết hợp) để khảo sát mối quan hệ
- Thông thường biến phụ thuộc được đặt theo hàng, biến độc lập được đặt theo cột
- Thực tế theo thói quen biến ít trị số được đặt trong cột
- Thông thường khi giả định rằng có sự tồn tại mối quan hệ giữa 2 biến thì để khảo sát bạn phải quyết định xem biến nào là biến độc lập và biến nào là biến lệ thuộc.
- Mở tệp Employee_data.sav có sẵn trong SPSS, chọn Analyze/ Descriptive Statistics/ Crosstabs, chọn biến phụ thuộc jobcat vào hộp Rows và chuyển biến độc lập minority vào hộp Column(s)/ OK.

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Employment Category * Minority Classification	474	100.0%	0	.0%	474	100.0%

Employment Category * Minority Classification Crosstabulation

Count		Minority Classification		Total
		No	Yes	
Employment Category	Clerical	276	87	363
	Custodial	14	13	27
	Manager	80	4	84
Total		370	104	474

* Thêm biến định tính thứ ba: Dùng làm biến điều khiển

SPSS cho phép bạn sử dụng thêm một biến định tính thứ ba làm điều khiển. Thí dụ, bạn nghi ngờ mối quan hệ giữa jobcat và minority là khác nhau đối với nam và nữ. Như vậy, bạn có thể tạo ra bảng chéo của jobcat và minority cho nam và nữ (biến gender). Chọn Analyze/Descriptive Statistics/Crosstabs, chuyển biến jobcat vào hộp Rows, minority vào hộp Columns và gender vào hộp thứ ba, nhấp Continue/OK.

Case Processing Summary

	Valid		Cases Missing		Total	
	N	Percent	N	Percent	N	Percent
Employment Category * Minority Classification * Gender	474	100.0%	0	.0%	474	100.0%

Employment Category * Minority Classification * Gender Crosstabulation

Count			Minority Classification		Total
Gender			No	Yes	
Female	Employment	Clerical	166	40	206
	Category	Manager	10	0	10
	Total		176	40	216
Male	Employment	Clerical	110	47	157
	Category	Custodial	14	13	27
		Manager	70	4	74
Total			194	64	258

* Kiểm định ý nghĩa của mối quan hệ mẫu

Muốn biết mối quan hệ trong mẫu có đủ mạnh để thuyết phục chúng ta rằng cũng có mối quan hệ tương tự trong tổng thể cần kiểm định thống kê với giả thiết rằng mẫu được chọn ngẫu nhiên từ tổng thể:

- Dùng kiểm định Chi-square để kiểm định mối quan hệ giữa 2 biến định tính
- Thông kê Chi-square không phải là số đo mức độ chặt chẽ của mối quan hệ, mà chỉ có ý nghĩa thống kê. Thống kê Chi - square chỉ phù hợp nếu đủ các dữ liệu.
- Theo kinh nghiệm, nếu có > 20% ô có tần số kỳ vọng nhỏ hơn 5 thì Chi-square là không thích hợp.

Tiếp tục sử dụng tệp Employee_data.sav, chọn Analyze/Descriptive Statistics/Crosstabs, nhấp nút Statistics, chọn Chi-square

Chi-Square Tests

Gender		Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Female	Pearson Chi-Square	2.383 ^b	1	.123	.214	.123
	Continuity Correction ^a	1.270	1	.260		
	Likelihood Ratio	4.205	1	.040		
	Fisher's Exact Test					
	Linear-by-Linear Association	2.372	1	.124		
	N of Valid Cases	216				
Male	Pearson Chi-Square	25.034 ^c	2	.000		
	Likelihood Ratio	28.905	2	.000		
	Linear-by-Linear Association	13.129	1	.000		
	N of Valid Cases	258				

a. Computed only for a 2x2 table

b. 1 cells (25.0%) have expected count less than 5. The minimum expected count is 1.85.

c. 0 cells (.0%) have expected count less than 5. The minimum expected count is 6.70.

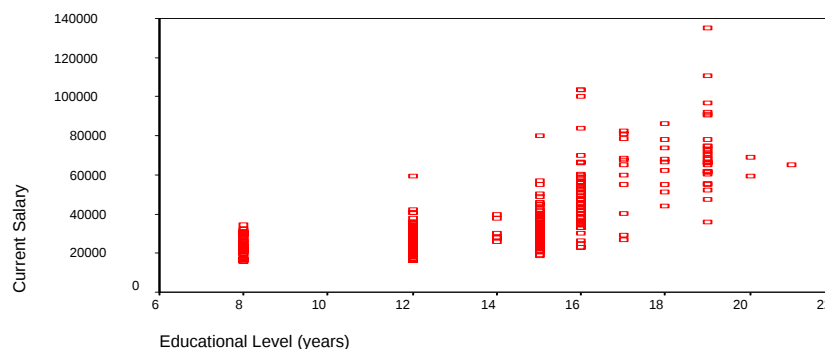
Bảng kết quả trình bày các giá trị Pearson Chi-square, Likelihood Ratio Chi-square, ... Quan hệ trong mẫu có ý nghĩa ở mức 0,05 nếu giá trị p (p-value) nhỏ hơn 0,05. Như vậy, trong kết quả trên thì quan hệ giữa jobcat và minority là độc lập (p=0.123) và sự lựa chọn tiêu thức phân tổ là hoàn toàn phù hợp.

b) Mối quan hệ giữa các biến định lượng

* Mô tả mối quan hệ giữa hai biến định lượng

- Dùng biểu đồ phân tán (XY - scatter) cho việc mô tả giữa hai biến định lượng

- Từ biểu đồ phát hiện các phần tử bất thường trong phân bố
- Trong quá trình mô tả mối quan hệ giữa hai biến định lượng cần quan sát 3 nội dung:
 - + Hướng (direction): Quan hệ thuận hay nghịch, ý nghĩa của mối quan hệ?
 - + Dạng (form): Dạng tuyến tính hay phi tuyến, sự phù hợp?
 - + Mức độ (strength): Dùng Regression và Descriptive để giải thích các thông số cần đo mức độ. Từ tệp Employee_data.save, chọn Graphs/Scatter/Simple/Define, chọn biến salary vào hộp Y Axis, chọn biến educ vào hộp X Axis/OK.



Ta thấy rằng mối quan hệ đồng biến giữa số năm học (educ) và mức lương. Thông qua việc xem xét mối quan hệ giữa các biến định lượng có cơ sở phân tích tương quan.

- Phân tích tương quan: Chọn Analyze/Corelate/Bivariate

Correlations

		Educational Level (years)	Current Salary
Educational Level (years)	Pearson Correlation	1	.661**
	Sig. (2-tailed)	.	.000
	N	474	474
Current Salary	Pearson Correlation	.661**	1
	Sig. (2-tailed)	.000	.
	N	474	474

** . Correlation is significant at the 0.01 level (2-tailed).

Hệ số tương quan giữa số năm học (educ) và mức lương (salary) là 0,661 có ý nghĩa thống kê ở mức 1%. Đây là cơ sở để tiếp tục xây dựng mô hình phân tích hồi quy.

* Thêm biến điều khiển định tính

- Để nghiên cứu sâu hơn mối quan hệ giữa các biến định lượng, người ta dùng thêm biến định tính nhằm kiểm tra cường độ tương quan giữa chúng thông qua các hệ số trong bảng ma trận.
- Chọn thêm biến định tính (minority) trong tệp Employee_data.sav kết hợp 2 biến định lượng educ và salary để phân tích ma trận tương quan bằng cách chọn Analyze/ Corelate/ Bivariate:

Correlations

		Educational Level (years)	Current Salary	Minority Classification
Educational Level (years)	Pearson Correlation	1	.661**	-.133**
	Sig. (2-tailed)	.	.000	.004
	N	474	474	474
Current Salary	Pearson Correlation	.661**	1	-.177**
	Sig. (2-tailed)	.000	.	.000
	N	474	474	474
Minority Classification	Pearson Correlation	-.133**	-.177**	1
	Sig. (2-tailed)	.004	.000	.
	N	474	474	474

** . Correlation is significant at the 0.01 level (2-tailed).

Ta thấy rằng giữa 2 biến định lượng (educ và salary) có mối quan hệ đồng biến. Khi thêm biến định tính minority vào ma trận tương quan cho thấy mối quan hệ giữa nó với biến định lượng educ là nghịch biến. Điều đó có nghĩa mối quan hệ giữa minority và salary cũng nghịch biến. Cả 2 hệ số tương quan này đều có ý nghĩa thống kê ở mức 1%.

c) Mối quan hệ giữa biến định lượng và biến định tính

* Quan hệ giữa một biến định lượng và một biến định tính

Mở tệp Employee_data.sav, khảo sát mối quan hệ giữa biến định lượng là salary và biến định tính gender. Cần xác định biến nào làm yếu tố ảnh hưởng. Trong thí dụ này ta xem yếu tố giới là nam hay nữ ảnh hưởng đến mức lương.

Chọn Analyze/Descriptive Statistics/Explore, chuyển salary vào hộp Dependent List và gender vào hộp Factor List. Thao tác này sẽ tạo ra thống kê tóm lược và biểu đồ thân và lá, biểu đồ hộp cũng được tạo ra rất thuận lợi cho việc so sánh.

Case Processing Summary

	Gender	Cases					
		Valid		Missing		Total	
		N	Percent	N	Percent	N	Percent
Current Salary	Female	216	100.0%	0	.0%	216	100.0%
	Male	258	100.0%	0	.0%	258	100.0%

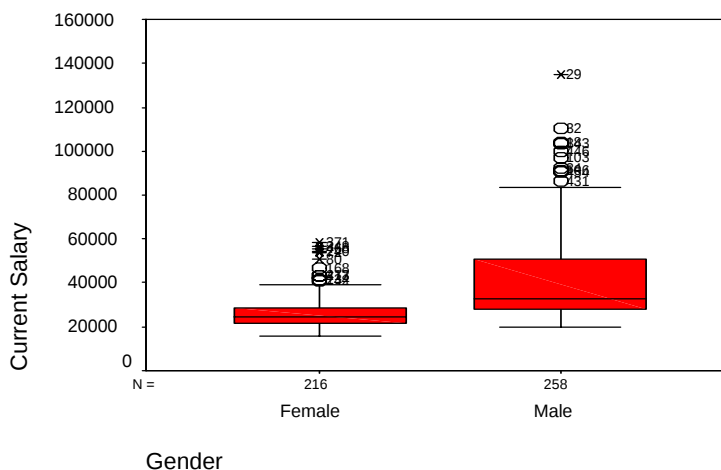
Current Salary Stem-and-Leaf Plot for
GENDER= Female

```

Frequency      Stem & Leaf
  2.00          1 . 55
 16.00          1 . 6666666666777777
 14.00          1 . 888899999999999
 31.00          2 . 000000000000001111111111111111
 35.00          2 . 22222222222222222222333333333333
 38.00          2 . 44444444444444444444444444445555555555
 22.00          2 . 666666666666777777777777777
 17.00          2 . 8888889999999999999
   7.00          3 . 0001111
   8.00          3 . 22233333
   8.00          3 . 44444555
   5.00          3 . 66777
   2.00          3 . 88
 11.00 Extremes      (>=40800)
  
```

Stem width: 10000
Each leaf: 1 case(s)

Descriptives					
Gender		Statistic		Std. Error	
Current Salary	Female	Mean	*****	\$514.258	
		95% Confidence Interval for Mean	*****		
		Lower Bound	*****		
		Upper Bound	*****		
		5% Trimmed Mean	*****		
		Median	*****		
		Variance	5.7E+07		
		Std. Deviation	*****		
		Minimum	\$15,750		
		Maximum	\$58,125		
		Range	\$42,375		
		Interquartile Range	\$7,012.50		
		Skewness	1.863	.166	
		Kurtosis	4.641	.330	
	Male	Mean	*****	*****	
		95% Confidence Interval for Mean	*****		
		Lower Bound	*****		
		Upper Bound	*****		
		5% Trimmed Mean	*****		
		Median	*****		
		Variance	3.8E+08		
		Std. Deviation	*****		
		Minimum	\$19,650		
		Maximum	\$135,000		
		Range	\$115,350		
		Interquartile Range	*****		
		Skewness	1.639	.152	
		Kurtosis	2.780	.302	



* Quan hệ giữa một biến định lượng và hai biến định tính

Giả sử bạn tìm thấy rằng có mối quan hệ giữa số năm học (educ) với giới tính (gender), nhưng liệu mối quan hệ này có giống nhau giữa những người thuộc nhóm dân tộc (minority) khác nhau. Vì vậy, bạn thêm vào biến định tính gender làm biến điều khiển để phân tích.

Chọn Analyze/Compare Means/Means, chuyển educ vào hộp Dependent List và minority vào hộp Independent List, nhấp Next, chuyển gender vào lớp thứ hai (Layer 2 ở 2) của Independent List, nhấp OK.

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
Educational Level (years) * Minority Classification * Gender	474	100.0%	0	.0%	474	100.0%

Report

Educational Level (years)				
Minority Classification	Gender	Mean	N	Std. Deviation
No	Female	12.34	176	2.407
	Male	14.92	194	2.848
	Total	13.69	370	2.942
Yes	Female	12.50	40	1.908
	Male	12.94	64	2.889
	Total	12.77	104	2.555
Total	Female	12.37	216	2.319
	Male	14.43	258	2.979
	Total	13.49	474	2.885

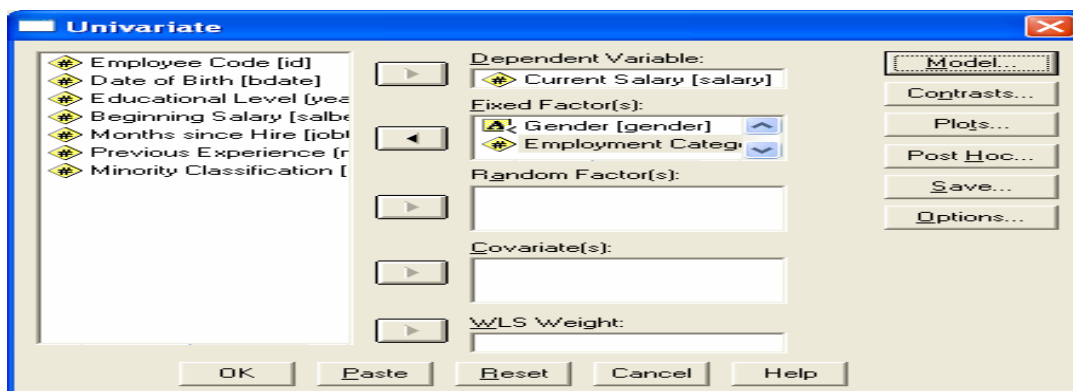
Bạn xem kết quả về số năm học bình quân của đối tượng nam và nữ theo thành phần dân tộc.

2. Phân tích phương sai

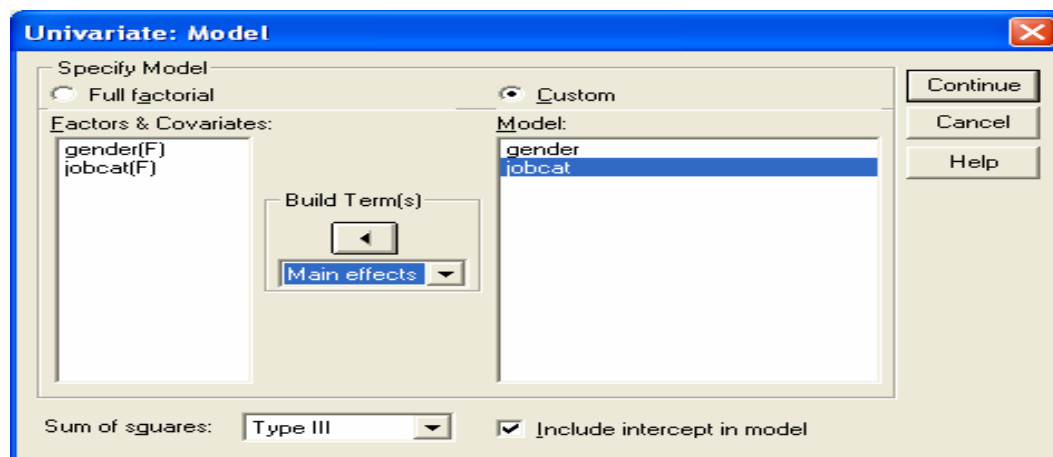
2.1 Phân tích phương sai đa nhân tố lên một biến phụ thuộc

Phương pháp này dùng để nghiên cứu tác động đồng thời của nhiều nhân tố lên một biến định lượng là biến phụ thuộc. Các nhân tố gồm những biến định tính có thứ bậc hoặc không có thứ bậc, có kiểu số hoặc kiểu chuỗi ký tự ngắn (không quá 8 ký tự), làm vai trò phân nhóm các quan sát. Các hiệp biến là những biến định lượng có quan hệ với biến phụ thuộc.

Mở tệp Employee_data.sav, thực hiện các thủ tục phân tích: Analyze/General Linear Model/Univariate...



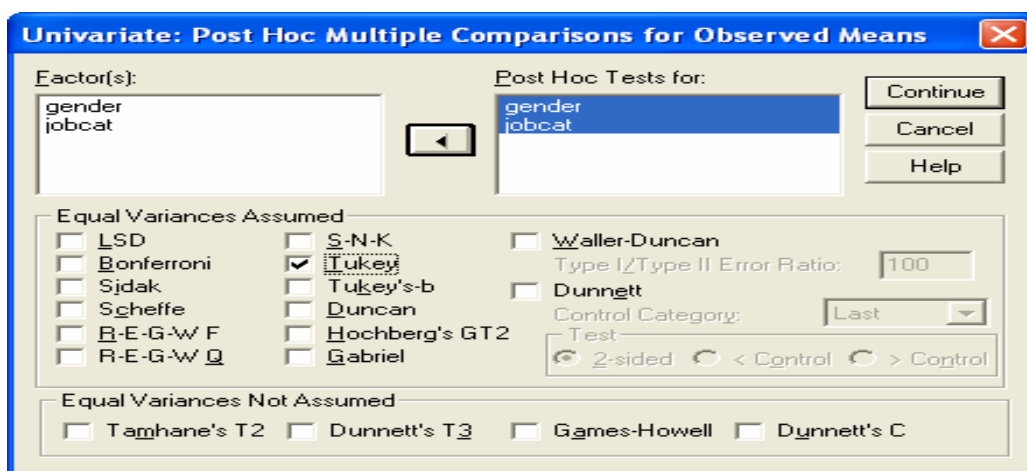
- Có thể đưa một biến định lượng từ danh sách biến vào ô WLS Weight để gán trọng số cho các quan sát, nếu các quan sát không có vai trò hoàn toàn như nhau.
- Nhấn nút Model để có thể quy định cho thủ tục đi sâu phân tích theo những mô hình cụ thể:



Các lựa chọn trên hộp thoại đó có công dụng như sau:

- Khung Specify Model có 2 nút:

- + Tích vào nút Full Factorial, việc phân tích sẽ cho ra kết quả với hiệu quả chính của tất cả mỗi nhân tố cùng hiệu quả tương tác của từng cặp nhân tố. Lựa chọn này không đưa ra hiệu quả tương tác của các hiệp biến.
- + Nút Custom được tích, ta có thể thao tác với các lựa chọn trong khung Build Term(s) để tự ấn định việc phân tích với các hiệu quả chính của một số nhân tố, hiệp biến cùng hiệu quả của nhóm các tương tác giữa các nhân tố với nhau và với cả hiệp biến.
- Nhấn vào phím mũi tên trong ô Interaction, ta có thể cho hiện lên bảng danh sách để ấn định cho tính ra các loại tương tác do các nhân tố và các hiệp biến trên phối hợp tạo thành.
- Ô Sum of squares để ấn định phương pháp tính tổng bình phương dùng đánh giá các giả thiết về sự khác biệt của các trung bình.
- Có thể tích hoặc cởi tích với ô Include Intercept in Model để đưa hằng số hồi quy vào mô hình hoặc loại bỏ hằng số đó khỏi mô hình (đối với trường hợp có các hiệp biến).
- Nhấn nút Continue, chọn tiếp Constrast ... để dùng các hệ số tương phản kiểm tra các sai khác giữa các mức của một nhân tố.
- Nhấn nút Plots ... để ấn định các loại đồ thị cần biểu diễn nhằm so sánh biến động của trị trung bình theo các nhân tố khác nhau.
- Có thể nhấn nút Post Hoc để ấn định các phép kiểm định hậu nghiệm cần tiến hành.



- Có thể nhấn nút Options để lựa chọn tính ra các tham số thống kê của mô hình hiệu quả cố định, kết quả cụ thể như sau:

Between-Subjects Factors

		Value Label	N
Gender	f	Female	216
	m	Male	258
Employment Category	1	Clerical	363
	2	Custodial	27
	3	Manager	84

Tests of Between-Subjects Effects

Dependent Variable: Current Salary

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	9.521E+10 ^a	3	3.174E+10	349.257	.000
Intercept	2.322E+11	1	2.322E+11	2555.614	.000
GENDER	5770190492	1	5770190492	63.501	.000
JOB CAT	6.729E+10	2	3.365E+10	370.264	.000
Error	4.271E+10	470	90867704.29		
Total	6.995E+11	474			
Corrected Total	1.379E+11	473			

a. R Squared = .690 (Adjusted R Squared = .688)

Loại III (Type III): Phương pháp này hay được dùng nhất nên được mặc định trong chương trình. Với phương pháp này, tổng bình phương của một hiệu quả được tính toán hiệu

chỉnh theo hiệu quả khác không bao hàm nó và lại trực giao với bất cứ hiệu quả nào bao hàm nó. Các tổng bình phương loại III có ưu điểm chính là nó bất biến đối với tần số trong các ô, chừng nào dạng ước lượng tổng quát còn được giữ nguyên. Vì thế phương pháp này rất hữu ích cho các mô hình không cân bằng (số quan sát trong các ô không đều nhau) không có các ô trống. Đối với các thiết kế nhân tố không có ô trống, phương pháp này tương đương với kỹ thuật trung bình trọng số bình phương của Yate. Phương pháp loại III này thường dùng cho:

- Một trong số các mô hình loại I và loại II:

+ Loại I (Type I): Phương pháp này còn gọi là phương pháp phân tích thứ bậc các tổng bình phương. Trong mô hình, mỗi thành phần được hiệu chỉnh theo duy nhất một thành phần khác xếp trước nó. Đây là phương pháp thường được dùng cho mô hình phân tích phương sai cân bằng (tức là tất cả các ô có số quan sát đều nhau), mô hình hồi quy đa thức (mỗi số hạng bậc thấp đều được xác định trước mọi số hạng bậc cao hơn) và mô hình thuần nhất (hiệu quả xác định bậc nhất được xác định hướng tới các hiệu quả xác định bậc hai, mỗi hiệu quả xác định bậc hai được định hướng tới các hiệu quả xác định bậc ba, v.v...).

+ Loại II (Type II): Theo phương pháp này thì tổng bình phương của một hiệu quả được hiệu chỉnh theo các hiệu quả thích hợp khác. Hiệu quả thích hợp là hiệu quả đáp ứng tất cả những hiệu quả không bao hàm hiệu quả đang xét. Phương pháp này thường được dùng cho phân tích phương sai cân bằng, mô hình chỉ xét các hiệu quả chính của nhân tố, mô hình hồi quy nào đó và mô hình lưới thuần nhất.

- Mô hình phân tích phương sai bất kỳ không chứa ô trống.

Kết quả phân tích phương sai trong ví dụ trên cho thấy nhân tố Gender và nhân tố Jobcat đều tác động đến biến phụ thuộc định lượng Current Salary. Mô hình phân tích phương sai đa nhân tố lên một biến phụ thuộc là định lượng có ý nghĩa thống kê ở mức 1%.

2.2. Phân tích phương sai đa nhân tố lên nhiều biến phụ thuộc

Thủ tục này cung cấp các phương pháp phân tích hồi quy và phân tích phương sai đối với nhiều biến phụ thuộc trong mối liên hệ với nhiều nhân tố và nhiều hiệp biến. Các nhân tố chia các quan sát thành nhiều nhóm. Sử dụng mô hình hồi quy tuyến tính tổng quát, ta có thể kiểm định giả thiết không về hiệu quả tác động của các nhân tố lên phân bố kết hợp của các biến độc lập. ta có thể chỉ rõ được hiệu quả riêng rẽ của từng cặp nhân tố cũng như hiệu quả tương tác của các nhân tố. Hơn nữa ta còn có thể phân tích cả hiệu quả của các hiệp biến và hiệu quả của tương tác với các hiệp biến. Đối với phân tích hồi quy, các biến độc lập được coi là các hiệp biến.

Đối với phân tích phương sai đa nhân tố lên nhiều biến phụ thuộc yêu cầu số liệu đưa vào phải thỏa mãn các điều kiện ràng buộc sau đây:

- Các biến phụ thuộc là các biến định lượng. Các nhân tố gồm những biến định tính có thứ bậc hoặc không có thứ bậc, có kiểu số hoặc kiểu chuỗi ký tự ngắn (không quá 8 ký tự), làm vai trò phân nhóm các quan sát. Các hiệp biến là những biến định lượng có quan hệ với biến phụ thuộc.

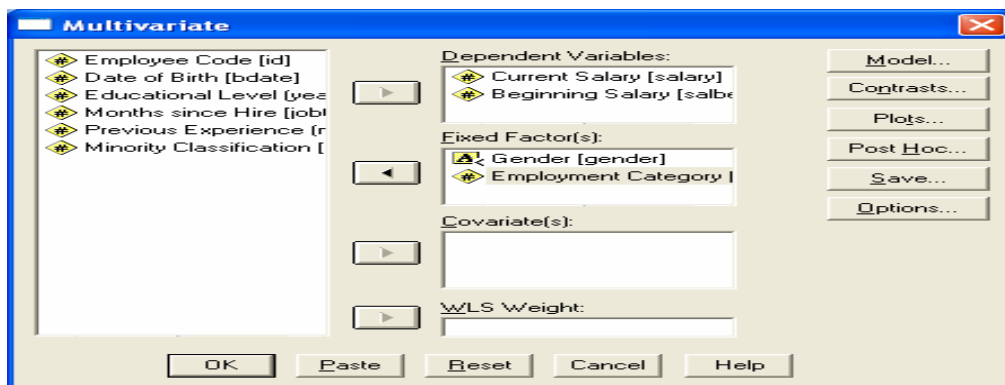
- Đối với các biến phụ thuộc, số liệu là các véc tơ được rút ngẫu nhiên ra từ quần thể có phân bố chuẩn nhiều chiều, với ma trận hiệp phương sai trong các ô của mảng nhiều chiều (do các nhân tố xác định) là như nhau. Nếu số quan sát trong mỗi ô là đủ lớn (40 trở lên), giả thiết phân bố chuẩn có thể bỏ qua, song vẫn nên đảm bảo tính đối xứng của số liệu trong từng nhóm. Để kiểm tra tính phù hợp của số liệu đối với mô hình, ta có thể kiểm định tính thuần nhất phương sai và vẽ các đồ thị của biến phụ thuộc dọc theo các mức của nhân tố. Nếu số liệu là kết quả của các phép đo lặp lại nhiều lần trên các đối tượng, nên dùng thủ tục của mô hình tuyến tính tổng quát với các phép đo lặp.

Để đánh giá và kiểm định các giả thiết có thể dùng các phương pháp lấy tổng bình phương loại I, loại II, loại II (mặc định) và loại IV (Type IV). Phương pháp lấy tổng bình phương loại IV được thiết kế cho trường hợp khi có các ô trống trong mô hình. Nếu một hiệu

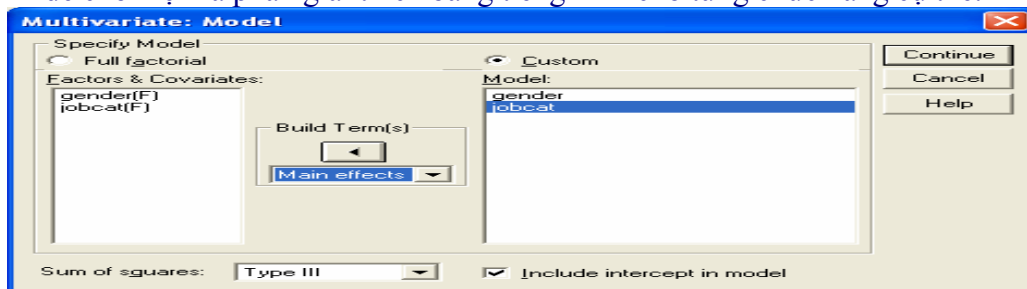
quả nào đó không chứa ô trống thì đối với nó, phương pháp này tương đương với các phương pháp loại II và loại III. Phương pháp này thường được dùng với một trong số các mô hình loại I và loại II, một mô hình phân tích phương sai có chứa ô trống.

Thủ tục đưa ra các thống kê mô tả cho tất cả các biến ở tất cả các ô trong mảng xác định bởi các nhân tố như các giá trị trung bình, độ lệch tiêu chuẩn và số lượng các quan sát trong mỗi ô. Đồng thời, thủ tục còn tính các tham số nhằm thực hiện phép kiểm định Levene về tính thuần nhất của phương sai. Các phép so sánh liên hợp và các tiêu chuẩn kiểm định hậu nghiệm do thủ tục cung cấp bao gồm các tiêu chuẩn kiểm định sai khác ý nghĩa nhỏ nhất, Bonferroni, Sidak, Scheffe, Student-Newman-Kuels, Duncan, Gabriel, Games-Howell, Dunnett (một phía và hai phía), tiêu chuẩn F bội Ryan-Einot-Gabriel-Welsch, thứ bậc bội Ryan-Einot-Gabriel-Welsch, sai khác ý nghĩa thực chất Tukey, tiêu chuẩn b-Tukey, GT2-Hochberg, tiêu chuẩn t-Waller-Duncan, T2-Tahane, T3-Dunnett và C-Dunnett. Có thể biểu diễn đồ thị giá trị trung bình của các biến phụ thuộc theo các mức của nhân tố, các giá trị phần dư và biểu đồ tương tác của các nhân tố.

Mở tệp Employee_data.sav, thực hiện các thủ tục phân tích: Analyze/General Linear Model/Multivariate...



Với cách sử dụng hoàn toàn tương tự, chỉ khác là ta sẽ đưa nhiều biến định lượng từ khung danh sách biến phía bên trái vào khung đầu tiên bên phải để làm các biến phụ thuộc, đồng thời cần phải xác định các nhân tố tác động ngẫu nhiên (không có khung Random Factor(s)). Các thao tác còn lại cũng được tiến hành giống như đối với thủ tục dành cho trường hợp một biến phụ thuộc đã được trình bày phía trên, với một vài điểm khác biệt nhỏ mà người dùng nếu quan tâm thì có thể tra cứu kỹ hơn bằng cách dùng phím Help trong các cửa hình để cho hiện ra phần giải thích bằng tiếng Anh cho từng chức năng cụ thể.



Kết quả cụ thể kết xuất như sau:

Between-Subjects Factors			
		Value Label	N
Gender	f	Female	216
	m	Male	258
Employment Category	1	Clerical	363
	2	Custodial	27
	3	Manager	84

Multivariate Tests ^d									
Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power ^a
Intercept	Pillai's Trace	.870	1566.820 ^b	2.000	469.000	.000	.870	3133.640	1.000
	Wilks' Lambda	.130	1566.820 ^b	2.000	469.000	.000	.870	3133.640	1.000
	Hotelling's Trace	6.682	1566.820 ^b	2.000	469.000	.000	.870	3133.640	1.000
	Roy's Largest Root	6.682	1566.820 ^b	2.000	469.000	.000	.870	3133.640	1.000
GENDER	Pillai's Trace	.148	40.689 ^b	2.000	469.000	.000	.148	81.378	1.000
	Wilks' Lambda	.852	40.689 ^b	2.000	469.000	.000	.148	81.378	1.000
	Hotelling's Trace	.174	40.689 ^b	2.000	469.000	.000	.148	81.378	1.000
	Roy's Largest Root	.174	40.689 ^b	2.000	469.000	.000	.148	81.378	1.000
JOBCAT	Pillai's Trace	.644	111.553	4.000	940.000	.000	.322	446.212	1.000
	Wilks' Lambda	.357	157.887 ^b	4.000	938.000	.000	.402	631.549	1.000
	Hotelling's Trace	1.797	210.281	4.000	936.000	.000	.473	841.125	1.000
	Roy's Largest Root	1.796	422.015 ^c	2.000	470.000	.000	.642	844.031	1.000

a. Computed using alpha = .05

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

d. Design: Intercept+GENDER+JOBCAT

Tests of Between-Subjects Effects									
Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power ^a
Corrected Model	Current Salary	9.521E+10 ^b	3	3.174E+10	349.257	.000	.690	1047.772	1.000
	Beginning Salary	1.939E+10 ^c	3	6462934194	306.452	.000	.662	919.355	1.000
Intercept	Current Salary	2.322E+11	1	2.322E+11	2555.614	.000	.845	2555.614	1.000
	Beginning Salary	5.420E+10	1	5.420E+10	2569.938	.000	.845	2569.938	1.000
GENDER	Current Salary	5770190492	1	5770190492	63.501	.000	.119	63.501	1.000
	Beginning Salary	1463258051	1	1463258051	69.383	.000	.129	69.383	1.000
JOBCAT	Current Salary	6.729E+10	2	3.365E+10	370.264	.000	.612	740.529	1.000
	Beginning Salary	1.328E+10	2	6639010902	314.801	.000	.573	629.601	1.000
Error	Current Salary	4.271E+10	470	90867704.29					
	Beginning Salary	9912102382	470	21089579.54					
Total	Current Salary	6.995E+11	474						
	Beginning Salary	1.665E+11	474						
Corrected Total	Current Salary	1.379E+11	473						
	Beginning Salary	2.930E+10	473						

a. Computed using alpha = .05

b. R Squared = .690 (Adjusted R Squared = .688)

c. R Squared = .662 (Adjusted R Squared = .660)

Kết quả phân tích phương sai trong ví dụ trên cho thấy nhân tố Gender và nhân tố Jobcat đều tác động đến 2 biến phụ thuộc định lượng Current Salary và Beginning Salary. Mô hình phân tích phương sai đa nhân tố lên 2 biến phụ thuộc thuộc là định lượng có ý nghĩa thống kê ở mức 1%.

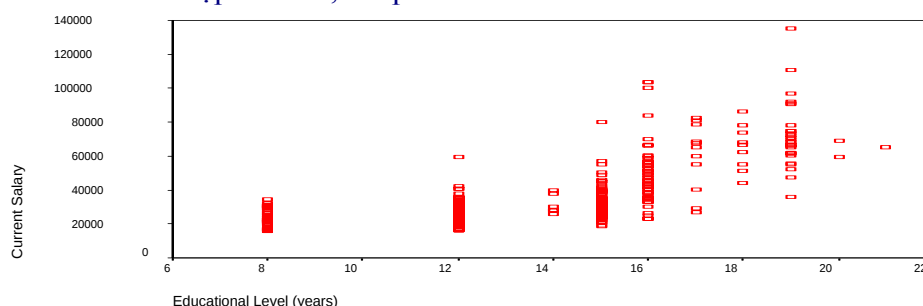
3. Phân tích hồi quy

3.1. Hồi quy tuyến tính đơn (Simple Linear Regression)

Hồi quy tuyến tính còn được gọi là OLS Regression (Ordinary Least Squares Regression), tức phương pháp làm cho bình phương các sai số (desiduals) được nhỏ nhất. Thủ tục hồi quy được sử dụng để khảo sát mối quan hệ tương quan giữa các biến.

Trước khi thực hiện việc phân tích mối quan hệ của hai biến định lượng bạn nên quan sát biểu đồ phân tán (Scatter) của chúng. Theo quy ước, chúng ta thường đặt biến phụ thuộc trên trục tung và biến độc lập trên trục hoành.

Mở tệp Employee_data.sav, quan sát mối quan hệ giữa mức lương (salary) và số năm đi học (educ). Giả định salary là biến phụ thuộc và educ là biến độc lập. Để xem xét biểu đồ phân tán bạn thực hiện như sau: Chọn Graphs/Scatter/Simple/Define, chuyển salary vào hộp Y Axis và educ vào hộp X Axis, nhấp OK.

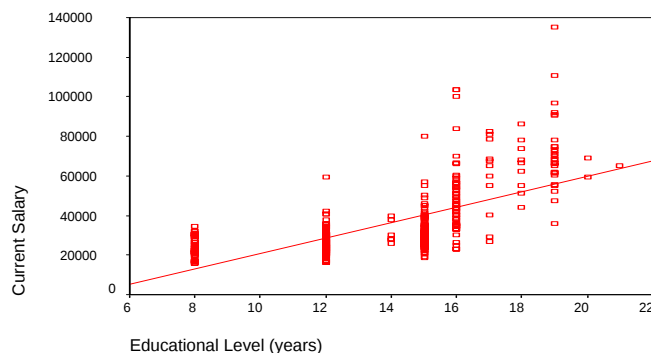


Trong biểu đồ phân tán bạn để ý xem các phần tử biến động thất thường. Trong quá trình mô tả mối quan hệ giữa 2 biến, bạn cần quan sát 3 nội dung là hướng (Direction), dạng (Form) và mức độ (Strength).

- Hướng: Bạn cần chọn phát biểu hợp lý nhất cho mỗi quan hệ đó là thuận hay nghịch, hay không có tính xu hướng, v.v...

- Dạng: Bạn nên xem xét quan hệ ở dạng tuyến tính hay phi tuyến. Trong trường hợp nghiên cứu này là quan hệ tuyến tính, như vậy đường thẳng biểu diễn tốt nhất mối quan hệ.

Nhấp đúp trên đồ thị chọn Chart/Options/Nhấp Total trong hộp Fit Line/OK.



- Mức độ: Bạn cần một thước đo định lượng về mức độ quan hệ và phương trình mô tả hồi quy, chọn Analyze/Regression/Linear, chuyển salary vào hộp Dependent và educ vào hộp Independents, để nhận được hệ số tương quan bạn nhấp Statistics và chọn Descriptives, có thể chọn thêm Estimates, Confidence intervals, Model fit., R Squared change .../OK

Correlations			
		Current Salary	Educational Level (years)
Pearson Correlation	Current Salary	1.000	.661
	Educational Level (years)	.661	1.000
Sig. (1-tailed)	Current Salary	.	.000
	Educational Level (years)	.000	.
N	Current Salary	474	474
	Educational Level (years)	474	474

Hệ số tương quan Pearson r chỉ đo chính xác mức độ quan hệ nếu cả hai biến có dạng số liệu liên tục và nếu quan hệ là tuyến tính. Nó đo lường mức độ phân tán quanh đường thẳng phù hợp nhất với số liệu. Nếu quan hệ phi tuyến hệ số tương quan Pearson r sẽ cho giá trị thấp trong mối quan hệ. Bạn nên sử dụng biểu đồ phân tán để kiểm định tính phi tuyến trước khi sử dụng Pearson r để tóm lược mức độ quan hệ.

Kết quả trên cho thấy tương quan giữa mức lương và số năm học là có ý nghĩa thống kê ($r = 0,661$, $n = 474$, $p < 0,0005$).

Mức ý nghĩa 1 phía (Sig.(1-tailed)) được tính dựa theo mối quan hệ một chiều (tương quan thuận trong trường hợp này) giữa số năm học và mức lương.

Variables Entered/Removed ^b			
Model	Variables Entered	Variables Removed	Method
1	Educational Level (years) ^a	.	Enter

a. All requested variables entered.

b. Dependent Variable: Current Salary

Model Summary									
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.661 ^a	.436	.435	\$12,833.540	.436	365.381	1	472	.000

a. Predictors: (Constant), Educational Level (years)

- R là hệ số tương quan biểu diễn cường độ của mối quan hệ giữa biến độc lập và biến phụ thuộc trong mô hình.

- R square là hệ số xác định R^2 (Coefficient of Determination), là giá trị bình phương của hệ số Pearson r. Hệ số xác định R^2 là tỷ số giữa ES/TSS, đo độ biến động trong biến phụ thuộc (salary) được giải thích bởi các biến động trong biến độc lập (educ). Trong kết quả đã tính $R^2 = 0,436$, mô hình hồi quy tuyến tính của chúng ta có thể giải thích được 43,6% sự khác biệt về mức lương giữa các đối tượng. R^2 là thước đo sự phù hợp của mô hình hồi quy tuyến tính đối với tổng thể, khi giá trị của nó bằng 0 thì không phải lúc nào cũng có nghĩa là giữa hai biến không có liên hệ gì cả, mà nó chỉ nói lên rằng hai biến không có liên hệ tuyến tính.

- Adjusted R square là R^2 điều chỉnh, được sử dụng để làm cho hệ số xác định R^2 phản ánh sát hơn sự phù hợp của mô hình với tổng thể. Hệ số này dùng để đo phần phương sai (variance) trong biến phụ thuộc (salary) được giải thích bởi biến động trong biến độc lập (educ). Adjusted R-square được tính như sau: $R_a^2 = R^2 - [(k*(1-R^2))/(N-k-1)]$

Với R^2 là hệ số xác định, k là số biến độc lập trong mô hình, N là số cặp quan sát.

Trong ví dụ này $R_a^2 = 0,435$, cho thấy có 43,5% biến động mức lương được giải thích bởi số năm học.

- Std. Error of the Estimate: Sai số chuẩn, đo sự phân tán của các biến lệ thuộc ước lượng chung quanh trung bình của nó. Nếu sai số chuẩn này lớn hơn 10% giá trị trung bình dự đoán của biến phụ thuộc thì được xem là cao.

- R^2 Change là một trong những cách đánh giá tầm quan trọng của các biến độc lập là xem xét mức độ tăng của R^2 khi một biến độc lập được đưa vào mô hình khi mô hình đã chứa sẵn các biến độc lập khác. Mức độ tăng này là $R^2 \text{ Change} = R^2 - R^2(i)$

Trong đó $R^2(i)$ là bình phương hệ số tương quan bội khi tất cả các biến độc lập đang ở trong mô hình ngoại trừ biến thứ i.

Căn bậc hai của $R^2 \text{ Change}$ là hệ số tương quan từng phần (part correlation coefficient). Trong bảng trên giá trị $R^2 \text{ Change}$ giống như R^2 (0,436) vì ta chỉ có một biến độc lập trong phương trình.

- F Change dùng kiểm định giả thiết H_0 là $R^2 \text{ Change}$ trong tổng thể bằng 0, còn được gọi là kiểm định F riêng (partial F test). $F \text{ Change} = R^2 \text{ Change}(N-k-1)/m(1-R^2)$

Trong đó m là số biến được đưa vào mô hình trong bước này, F Change được tính mức ý nghĩa từ phân phối F với k và N-k-1 bậc tự do.

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6.02E+10	1	6.018E+10	365.381	.000 ^a
	Residual	7.77E+10	472	164699740.8		
	Total	1.38E+11	473			

a. Predictors: (Constant), Educational Level (years)

b. Dependent Variable: Current Salary

Ý nghĩa của bảng ANOVA là để kiểm định xem hệ số R khác 0 có ý nghĩa hay không. Giá trị $F = 365,381$, $df = 1$, $p < 0,0005$ cho thấy mối quan hệ tuyến tính có ý nghĩa giữa chiều cao và cân nặng.

Giá trị Total Sum of Squares (TSS) là tổng độ lệch trong biến lệ thuộc. Mục đích của hồi quy là giải thích các độ lệch này bằng cách tìm giá trị beta tốt nhất có thể nhỏ nhất tổng bình phương các độ lệch này.

Giá trị Total Sum of Squares of Regression (ESS) còn gọi là Explained Sum of Squares là phần độ lệch từ TSS có thể giải thích được mô hình.

Giá trị Total Sum of Squares of Residual (RSS) là phần độ lệch không thể giải thích được bởi mô hình: $RSS = TSS - ESS$

Coefficients ^a								
		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-18331.2	2821.912		-6.496	.000	-23876.242	-12786.114
	Educational Level (years)	3909.907	204.547	.661	19.115	.000	3507.971	4311.842

a. Dependent Variable: Current Salary

- Beta là hệ số hồi quy chuẩn hoá (Standard regression coefficient). $Beta = b(S_X/S_Y)$
 Với b là hệ số hồi quy, S_X là độ lệch chuẩn của biến độc lập và S_Y là độ lệch chuẩn của biến phụ thuộc.

Thực chất thì hệ số beta là độ dốc của đường thẳng (tìm được theo phương pháp bình phương bé nhất) khi cả 2 biến X và Y được biểu diễn bằng thang đo chuẩn hoá (Z-score).

- Giá trị b của biến educ là độ dốc (slope) của đường hồi quy.
- Giá trị b của Constant là giá trị cắt trục tung (intercept)
- Giá trị t trong bảng trên, đại lượng thống kê t được sử dụng để kiểm định giả thiết H_0 rằng hệ số hồi quy (độ dốc của tổng thể) bằng 0. Giá trị của $t = 19,115$, $p < 0,0005$ cho thấy giả thiết H_0 bị bác bỏ, tức có mối liên hệ tuyến tính thuận có ý nghĩa thống kê giữa mức lương và số năm học. Công thức : $t = b/Se(b)$ với b là hệ số hồi quy, $Se(b)$ là độ lệch chuẩn của b, phân phối t với N-2 bậc tự do.
- Confidence Interval (khoảng tin cậy): Một giá trị thống kê tính từ mẫu quan sát cho chúng ta một ước lượng điểm (point Estimation) về tham số chưa biết của tổng thể. Giá trị ước lượng từ mẫu thì thường khác với các giá trị tham số chưa biết của tổng thể nhưng ta hy vọng các giá trị này không quá chênh lệch nhau. Dựa trên ước lượng mẫu chúng ta có thể tính được phạm vi giá trị chứa giá trị tham số của tổng thể với một xác suất được xác định (95% hay 99%). Phạm vi giá trị này được gọi là khoảng tin cậy.

3.2 Hồi quy tuyến tính bội (Multiple Linear Regression)

Trong phần hồi quy bội này chúng ta sẽ nghiên cứu xem mức lương (salary) phụ thuộc vào số năm học (educ), số năm kinh nghiệm (prevexp), thành phần dân tộc (minority).

Chọn Analyze/Regression/Linear, chuyển biến salary vào hộp Dependent và các biến khác như educ, prevexp, minority vào hộp Independents, trong phần Method bạn có thể chọn 1 trong 5 phương pháp sau đây để xây dựng mô hình hồi quy:

- Enter (đưa vào một lượt): Đây là phương pháp mặc định của chương trình. Các biến trong khối sẽ được đưa vào mô hình cùng một lúc.
- Stepwise (từng bước): Đưa vào và rút ra các biến từng bước một. Các biến trong khối sẽ được xem xét lại từng bước một khi đưa vào hay rút ra khỏi mô hình.
- Remove (loại bỏ một lượt): các biến trong khối sẽ được rút ra chỉ trong một bước.
- Backward (rút ra dần): các biến trong khối sẽ được đưa vào mô hình cùng một lúc và sau đó được rút ra dần dựa vào tiêu chuẩn rút ra.
- Forward (đưa vào dần): Các biến trong khối sẽ được lần lượt đưa vào mô hình dựa vào tiêu chuẩn đưa vào.

Các biến phải vượt qua được tiêu chuẩn chấp nhận (Tolerance criterion) mới được đưa vào mô hình, cho dù ta có chỉ định phương pháp đưa biến vào đi nữa. Độ chấp nhận mặc định là 0,0001.

Một biến cũng không được đưa vào phương trình nếu nó làm cho độ chấp nhận của một biến đã được đưa vào mô hình xuống dưới tiêu chuẩn chấp nhận.

Đối với phương pháp Stepwise, số bước tối đa để đưa các biến vào phương trình là bằng 2 lần số biến độc lập. Đối với phương pháp Backward và Forward thì số bước tối đa bằng với số biến thoả mãn tiêu chuẩn đưa vào rút ra.

Tất cả các biến độc lập được chọn sẽ được đưa vào một mô hình hồi quy duy nhất. Tuy nhiên, bạn có thể chỉ định các phương pháp đưa biến vào khác nhau cho các tập hợp con

khác nhau của các biến độc lập. Thí dụ, bạn có thể đưa một khối biến vào mô hình hồi quy bằng phương pháp Forward và đưa khối biến thứ hai vào mô hình bằng phương pháp Stepwise. Để đưa khối biến thứ hai vào bạn nhấp chuột tại nút Next, tiếp tục như thế cho khối biến thứ ba chẳng hạn. Để di chuyển tới lui giữa các khối biến độc lập, bạn sử dụng các nút Previous và Next. Bạn có thể chỉ định đến 9 khối biến khác nhau.

- Trong hộp thoại Linear Regression, nhấp nút Statistics để mở hộp thoại Linear Regression: Statistics, bạn có thể chọn các tùy chọn:

- + Estimates (các ước lượng) gồm các hệ số hồi quy và các đo lường có liên quan.
- + Confidence intervals (khoảng tin cậy): Thể hiện khoảng tin cậy 95% của từng hệ số hồi quy không chuẩn hoá.
- + Covariance matrix (ma trận hiệp phương sai): Thể hiện ma trận hiệp phương sai của các hệ số hồi quy không chuẩn hoá.

Bạn cũng có thể chọn thêm một hay nhiều các thông số thống kê sau:

- + Descriptives (các thống kê mô tả): Các giá trị trung bình, độ lệch chuẩn và ma trận tương quan với các xác suất kiểm định một phía.
- + Model fit (các thống kê đánh giá sự phù hợp của mô hình): R , R^2 , R^2 điều chỉnh và sai số chuẩn. Ngoài ra bảng ANOVA sẽ thể hiện số bậc tự do, tổng các độ lệch bình phương, độ lệch phương sai, giá trị thống kê F và xác suất F quan sát được. Các thống kê đánh giá sự phù hợp của mô hình sẽ được thể hiện theo mặc định.
- + Durbin-Watson: Thể hiện thống kê kiểm định Durbin-Watson và cũng thể hiện các thống kê tóm tắt của ácc phần dư không chuẩn hoá và chuẩn hoá cũng như các giá trị dự đoán. Đại lượng này dùng để kiểm định tương quan của các sai số kế nhau, có giá trị $d = 0 - 4$. Nếu phần dư không có tương quan với nhau thì d sẽ gần 2. Giá trị $d < 2$ có nghĩa là ácc phần dư gần nhau có tương quan thuận. Giá trị $d > 2$ có nghĩa là các phần dư có liên hệ nghịch.
- + Collinearity Diagnostics (chẩn đoán hiện tượng đa cộng tuyến): Thể hiện qua hệ số phóng đại phương sai (Variation Inflation of Factor – VIF), các giá trị đặc trưng (Eigenvalues) của ma trận tích mô men chéo, các chỉ số điều kiện (Condition Index) và các tỷ lệ của các bộ phận phương sai. Tùy chọn này cũng sẽ thể hiện độ chấp nhận của các biến trong mô hình, của các biến không được đưa vào mô hình, độ chấp nhận của một biến nếu như nó được đưa vào trong mô hình ở bước tiếp theo.
- + Colinearity là đa cộng tuyến, phản ánh trạng thái của các biến độc lập có tương quan chặt chẽ với nhau, rất khó tách ảnh hưởng của từng biến một. Các công cụ chẩn đoán sau đây sẽ giúp ta phát hiện sự tồn tại của hiện tượng cộng tuyến:

Độ chấp nhận (Tolerance) của biến thường được sử dụng để đo lường hiện tượng cộng tuyến. Độ chấp nhận của biến i được định nghĩa là $1-R_i^2$, trong đó R_i là hệ số tương quan bội khi biến độc lập i được dự đoán từ các biến độc lập khác. Nếu độ chấp nhận của một biến nhỏ thì nó gần như là một kết hợp tuyến tính của ácc biến độc lập khác (tức có hiện tượng đa cộng tuyến).

Hệ số phóng đại phương sai có liên hệ gần gũi với độ chấp nhận. Thực tế nó là nghịch đảo của độ chấp nhận, tức là đối với biến i thì: $VIF = 1/(1-R_i^2)$. Gọi là hệ số phóng đại phương sai vì nó có liên quan đến việc tính toán phương sai của hệ số hồi quy của biến i . Khi VIF tăng thì phương sai của hệ số hồi quy cũng tăng theo.

- Để điều chỉnh các tiêu chuẩn biến vào hoặc ra khỏi mô hình hồi quy hay điều chỉnh việc xử lý các quan sát thiếu dữ kiện, bạn nhấp Option trong hộp thoại Linear Regression:

- + Phần Stepping Method Criteria: Tiêu chuẩn vào hay ra áp dụng cho các phương pháp đưa vào dần, loại trừ dần và từng bước. Bạn có thể chọn:

Use probability of F: Sử dụng xác suất F vào (Entry) và xác suất F ra (Removal) làm tiêu chuẩn đưa biến vào và ra. Đây là tiêu chuẩn mặc định. Giá trị vào mặc định là 0,05 và giá trị ra mặc định là 0,10. Muốn thay đổi các giá trị mặc định này, bạn nhập các xác suất F

mới vào. Cả hai giá trị phải lớn hơn 0 và nhỏ hơn hay bằng 1, và F vào (FIN) phải nhỏ hơn F ra (FOUT).

Use F value: Sử dụng các giá trị F làm tiêu chuẩn vào và ra. Giá trị vào mặc định là 3,84 và giá trị ra mặc định là 2,71.

+ Phần Missing values: Bạn có thể chọn một trong ba cách xử lý số liệu thiếu sau:

Exclude cases listwise: Chỉ có những quan sát nào có đầy đủ giá trị của tất cả các biến mới được đưa vào phân tích. Đây là cách xử lý mặc định.

Exclude cases pairwise: Các quan sát có đầy đủ dữ kiện đối với một cặp biến đang nghiên cứu liên hệ sẽ được sử dụng để tính hệ số tương quan làm cơ sở cho phân tích hồi quy. Bậc tự do được tính trên N nhỏ nhất của các cặp biến.

Replace with mean: Thay thế các dữ kiện thiếu bằng trung bình của biến và tất cả các quan sát đều được sử dụng để tính toán.

- Chọn Include Constant in Equation: Mô hình hồi quy có chứa hằng số. Đây là mặc định, tuy nhiên bạn có thể làm cho đường hồi quy đi qua gốc toạ độ bằng cách bỏ dấu chọn trong ô này.

- **Nhấp nút Plots để vẽ đồ thị phân tán của các biến phân tích.** Trong cửa sổ Linear Regression Plots bạn thấy biến phụ thuộc và các biến phần dư (Residual) và biến dự đoán (các mức độ lý thuyết của các biến phụ thuộc tính theo mô hình đã xây dựng) sẽ xuất hiện trong sách biến nguồn (Source List)

INDEPENDENT

* ZPRED: Giá trị dự đoán chuẩn hoá

* ZRESID: Phần dư chuẩn hoá

* DRESID: Phần dư loại bỏ của quan sát đang xem xét

* ADJPRED: Giá trị dự đoán điều chỉnh

* SRESID: Phần dư student hoá

* SDRESID: Phần dư loại bỏ của quan sát đang xem xét được student hoá

Hãy chọn một biến cho trục tung (y) và một biến cho trục hoành (x).

Để thực hiện được nhiều đồ thị, nhấp chuột tại nút next và lặp lại việc xác định biến cho trục tung và biến cho trục hoành. Bạn có thể chỉ định được 9 đồ thị cùng một lúc. Tất cả đồ thị đều được chuẩn hoá. Trong hộp Standardized Residual Plots (đồ thị phần dư chuẩn hoá) có các tùy chọn:

+ Histogram (biểu đồ tần số): Thể hiện biểu đồ tần số của các phần dư chuẩn hoá và đường cong chuẩn được đặt chồng lên biểu đồ này.

+ Normal Probability Plot (biểu đồ xác suất chuẩn): Đồ thị so sánh xác suất chuẩn của các phần dư chuẩn hoá với một phân phối chuẩn.

+ Produce all partial plots (biểu đồ phân tán từng phần): Vẽ biểu đồ phân tán của các phần dư của biến phụ thuộc và một biến độc lập khi cả hai biến này được hồi quy riêng. Các biểu đồ này sẽ thể hiện từng biến độc lập trong phương trình theo thứ tự giảm dần của sai số chuẩn của hệ số hồi quy, tất cả các biểu đồ này được chuẩn hoá. Tất nhiên phải có ít nhất hai biến độc lập trong phương trình thì ta mới có thể thực hiện được biểu đồ từng phần này.

- **Chọn Save trong hộp thoại Linear Regression** để mở hộp thoại Linear Regression Save. SPSS sẽ tự động gán tên biến mới cho bất cứ thông số nào bạn muốn sao lưu. Kết quả là bạn có thêm biến mới và nội dung của nó được ghi vào bảng chứa các biến nguồn trong hộp thoại Linear Regression.

+ Predicted Values (các giá trị dự đoán): Bạn có thể chọn một hoặc nhiều thông số sau:

* Unstandardized: Các giá trị dự đoán không chuẩn hoá.

Standardized: Các giá trị dự đoán chuẩn hoá.

Adjusted: Các giá trị dự đoán điều chỉnh.

S.E. of mean prediction: Sai số chuẩn của các giá trị dự đoán.

- * Distance (khoảng cách): Bạn có thể chọn một hay nhiều loại khoảng cách: Mahalanobis, Cook, Leverage values.
- * Prediction intervals (khoảng dự đoán): Bạn có thể chọn một hay cả hai loại dự đoán sau:
 Mean: Giới hạn trên và dưới của khoảng dự đoán của trị trung bình.
 Individual: Giới hạn trên và dưới của khoảng dự đoán cho từng quan sát.
- * Confidence interval (khoảng tin cậy): Khoảng tin cậy mặc định cho trị trung bình và từng quan sát là 95%. Bạn có thể thay đổi giá trị mặc định này với giá trị từ > 0 đến < 100.
- * Residuals (phần dư): Bạn có thể chọn một hay nhiều loại sau:
 Unstandardized: Phần dư không chuẩn hoá.
 Standardized: Phần dư chuẩn hoá
 Studentized: Phần dư Student hoá.
 Deleted: Phần dư loại bỏ quan sát đang xem xét.
 Studentized deleted: Phần dư loại bỏ quan sát đang xem xét được student hoá.
- * Influence Statistics (các thông số thống kê ảnh hưởng): Bạn có thể chọn một hoặc nhiều thông số sau:
 DfBetas phản ánh sự thay đổi hệ số hồi quy khi loại bỏ một quan sát nào đó ra khỏi quá trình tính toán SPSS sẽ tính lại tất cả các hệ số hồi quy trong phương trình, kể cả hằng số.
 Standardized DfBetas: Tính các Dfbeta chuẩn hoá
 DfFit phản ánh sự thay đổi của các giá trị dự đoán khi loại bỏ một quan sát đang xem xét ra khỏi quá trình tính toán.
 Standardized DfFit: Tính các DfFit chuẩn hoá.
 Covariance ratio: Tính tỷ số giữa các thành phần của ma trận phương sai, hiệp phương sai khi quan sát đang xem xét bị loại ra khỏi quá trình tính toán với các thành phần của ma trận phương sai, hiệp phương sai khi tất cả các quan sát được đưa vào tính toán. Nếu tỷ số này tiến đến 1 thì quan sát đang xem xét không ảnh hưởng nhiều đến ma trận hiệp phương sai.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.672 ^a	.451	.448	\$12,688.728	.451	128.868	3	470	.000
2	.000 ^b	.000	.000	\$17,075.661	-.451	128.868	3	470	.000

a. Predictors: (Constant), Minority Classification, Educational Level (years), Previous Experience (months)

b. Predictor: (constant)

ANOVA^c

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6.22E+10	3	2.075E+10	128.868	.000 ^a
	Residual	7.57E+10	470	161003826.0		
	Total	1.38E+11	473			
2	Regression	.000	0	.000	.	. ^b
	Residual	1.38E+11	473	291578214.5		
	Total	1.38E+11	473			

a. Predictors: (Constant), Minority Classification, Educational Level (years), Previous Experience (months)

b. Predictor: (constant)

c. Dependent Variable: Current Salary

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-19424.2	3109.426		-6.247	.000	-25534.331	-13314.136					
	Educational Level (years)	3958.883	210.070	.669	18.846	.000	3546.090	4371.675	.661	.656	.644	.927	1.079
	Previous Experience (months)	14.028	5.804	.086	2.417	.016	2.622	25.433	-.097	.111	.083	.924	1.083
	Minority Classification	-4158.559	1430.535	-.101	-2.907	.004	-6969.594	-1347.523	-.177	-.133	-.099	.969	1.032
2	(Constant)	34419.568	784.311		43.885	.000	32878.402	35960.733					

a. Dependent Variable: Current Salary

Coefficient Correlations^a

Model		Minority Classification	Educational Level (years)	Previous Experience (months)
1	Correlations	Minority Classification	1.000	.101
		Educational Level (years)	.101	1.000
		Previous Experience (months)	-.116	.238
	Covariances	Minority Classification	2046429.629	30244.863
		Educational Level (years)	30244.863	44129.369
		Previous Experience (months)	-962.825	289.832
				33.689

a. Dependent Variable: Current Salary

Collinearity Diagnostics^a

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions			
				(Constant)	Educational Level (years)	Previous Experience (months)	Minority Classification
1	1	2.842	1.000	.00	.00	.04	.04
	2	.690	2.029	.00	.00	.01	.93
	3	.449	2.516	.01	.01	.84	.01
	4	.019	12.212	.99	.98	.11	.02
2	1	1.000	1.000	1.00			

a. Dependent Variable: Current Salary

Excluded Variables^b

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics		
						Tolerance	VIF	Minimum Tolerance
2	Educational Level (years)	.661 ^a	19.115	.000	.661	1.000	1.000	1.000
	Previous Experience (months)	-.097 ^a	-2.128	.034	-.097	1.000	1.000	1.000
	Minority Classification	-.177 ^a	-3.915	.000	-.177	1.000	1.000	1.000

a. Predictor: (constant)

b. Dependent Variable: Current Salary

Trong ba thủ tục chọn biến như Stepwise, Backward, Forward (phần Method) thì không có thủ tục nào là tốt nhất, chúng chỉ đơn giản là nhận ra các biến có khả năng dự đoán “tốt nhất” biến phụ thuộc trong mẫu. Ba thủ tục này cũng không phải lúc nào cũng cho ra cùng một phương trình. Mô hình được chọn của bất kỳ phương pháp nào cũng nên được nghiên cứu cẩn thận xem có vi phạm giả thiết nào không. Chúng ta nên xây dựng nhiều mô hình có thể chấp nhận được và sau đó chọn ra một mô hình trên cơ sở khả năng giải thích, khả năng thu thập dữ liệu dễ dàng của biến.

Thông thường chúng ta không biết trước được mô hình tuyến tính có thích hợp hay không. Do đó, chúng ta cần tiến hành dò tìm trên các phần dư để xem coi có các bằng chứng về sự vi phạm giả thiết không.

Phần dư (residual): Khi xây dựng mô hình, phần dư là những biến thiên còn lại sau khi điều chỉnh theo mô hình. Chúng là chênh lệch giữa giá trị quan sát được và giá trị dự đoán (hay giá trị lý thuyết). $E_i = Y_i - \hat{Y}_i$

Khi phân tích hồi quy, sai số thực e_i được cho rằng có phân phối chuẩn với trung bình bằng 0 và phương sai không đổi σ^2 . Nếu mô hình phù hợp với dữ kiện thì phần dư quan sát được sẽ là ước lượng không chệch của e_i .

Phần dư có thể được điều chỉnh theo 2 phương pháp:

Phương pháp thứ nhất là chuẩn hoá bằng cách chia phần dư của quan sát thứ i cho độ lệch chuẩn của các phần dư trong mẫu quan sát. các phần dư chuẩn hoá có trung bình là 0 và độ lệch chuẩn là 1.

Phương pháp thứ hai là Student hoá (Studentize) phần dư bằng cách chia phần dư cho ước lượng độ lệch chuẩn thay đổi từ điểm này qua điểm khác theo khoảng cách từ X_i đến trung bình $\bar{X}$. Thông thường các phần dư chuẩn hoá và Student hoá có giá trị gần bằng nhau (nhưng điều này không luôn luôn đúng). Phần dư Student hoá phản ánh chính xác hơn khác biệt phương sai của sai số thực của các điểm quan sát.

Từ phương pháp Remove, kết quả tính ở trên cho thấy phương trình hồi quy được ước lượng cho thấy số năm học (educ), số năm kinh nghiệm (prevexp) và thành phần dân tộc (minority) là 3 biến dự báo tốt nhất cho mức lương (salary).

BÀI TẬP THỰC HÀNH CHƯƠNG III

Bài 1. Nghiên cứu về kết quả điểm số của các xạ thủ bắn súng phụ thuộc vào phương pháp ngắm bắn, với bảng số liệu sau:

Mở cả hai mắt	Chỉ mở mắt trái	Chỉ mở mắt phải
22	28	33
27	37	29
29	34	39
20	29	33
18	31	37
30	33	38

Dùng SPSS để phân tích phương sai 1 yếu tố (ANOVA – single factor)

Bài 2. Kết quả điểm số của 5 xạ thủ ở 3 phương pháp ngắm bắn như sau:

Xạ thủ	Phương pháp ngắm bắn		
	Mở cả hai mắt	Chỉ mở mắt trái	Chỉ mở mắt phải
Số 1	19	32	36
Số 2	29	33	37
Số 3	30	37	35
Số 4	32	34	39
Số 5	25	28	33

Dùng SPSS để phân tích phương sai theo khối với mẫu ngẫu nhiên theo khối (Randomized block design)

Bài 3. Một nghiên cứu được thực hiện nhằm xem xét sự liên hệ giữa loại phân bón, giống lúa và năng suất lúa (tạ/ ha), được ghi nhận từ các thực nghiệm sau:

Loại phân bón	Giống lúa								
	A			B			C		
1	65	68	62	69	71	67	75	75	78
2	74	79	76	72	69	69	70	69	65
3	64	72	65	68	73	75	78	82	80
4	83	82	84	78	78	75	76	77	75

Dùng SPSS để phân tích phương sai 2 yếu tố

Bài 4. Giả sử một hệ thống các siêu thị đang quan tâm đến hàng loạt các ảnh hưởng lên lượng hàng bán ra hàng tuần. Liệu chừng số lượng hàng bán của một sản phẩm cụ thể nào đó có phụ thuộc vào vị trí của giá/ kệ để hàng của sản phẩm này không? Nếu có, liệu chừng qui mô của gian hàng có phải là yếu tố đáng kể (có ý nghĩa) không? Và sự tương tác nào giữa vị trí giá để hàng và độ lớn của gian hàng không? Berenson và Levine (1992) đã nghiên cứu

vấn đề này, với cùng cỡ mẫu bằng nhau cho từng kết hợp giữa vị trí để giá hàng và độ lớn của gian hàng. Dữ liệu được thể hiện bảng sau (Lượng hàng bán hàng tuần theo vị trí giá để hàng và qui mô của cửa hàng)

Độ lớn gian hàng	Vị trí giá để hàng của sản phẩm			
	A	B	C	D
Nhỏ	45	56	65	48
	50	63	71	53
Trung bình	57	69	73	60
	65	78	80	57
Lớn	70	75	82	71
	78	82	89	75

Dùng SPSS để phân tích phương sai 2 yếu tố cùng cỡ mẫu.

Bài 5. Một thiết kế khối đầy đủ được ngẫu nhiên hoá với 2 yếu tố điều trị (Univariate ANOVA: A Randomized Complete Block Design with Two Treatments)

Độ lớn (khối lượng) của một ổ bánh mì bị ảnh hưởng của các kết hợp khác nhau của chất béo và chất hoạt tính bề mặt trong bột mì được nhào ra sao ? Sử dụng dữ liệu của Miliken và Johson (1984) sau đây:

Chất béo	Chất có hoạt tính bề mặt	Nguồn bột mì			
		1	2	3	4
1	1	6,7	4,3	5,7	-
	2	7,1	-	5,9	5,6
	3	-	5,5	6,4	5,8
2	1	-	5,9	7,4	7,1
	2	-	5,6	-	6,8
	3	6,4	5,1	6,2	6,3
3	1	7,1	5,9	-	-
	2	7,3	6,6	8,1	6,8
	3	-	7,5	9,1	-

Bột mì là thành phần cơ bản của bánh mì, nhưng sự khác nhau trong bột mì không phải là trọng tâm của thí nghiệm này. Các nhà sản xuất bánh mì đã nêu rằng sự khác nhau về khối lượng của bánh mì được gây ra bởi các chất béo và các chất có hoạt tính bề mặt lại có thể khác nhau giữa các lô bột khác nhau. Do vậy 4 lô bột mì cùng loại nhưng khác nguồn cung cấp được sử dụng để ngẫu nhiên hoá (randomize) những biến thiên không thể giải thích được từ các ảnh hưởng của 2 yếu tố điều trị giữa các nguồn bột mì. Để thiết lập một thiết kế khối đầy đủ (complete block design), tất cả các kết hợp của các yếu tố điều trị được chỉ định một cách ngẫu nhiên vào từng lô (nguồn) bột mì. Bởi vì mọi mức độ có thể có của từng yếu tố đều được thể hiện, đây là một mô hình các ảnh hưởng cố định (fixed-effects model)

Chú ý rằng có những ô trống trong bảng số liệu là do có một thùng bột ủ có men làm bánh không tốt lắm. tuy nhiên tất cả 9 kết hợp yếu tố điều trị (chất béo (fat) * chất có hoạt tính bề mặt (surfactant)) có thể có đều được quan sát ít nhất là một lần. Chính vì vậy, ước lượng của

tương tác fat*surfactant không bị ảnh hưởng của các ô rỗng. Trong tệp tin có 4 biến: fat (chất béo), surf(chất có hoạt tính bề mặt), flour (bột mì) và spvol (khối lượng ổ bánh). Dùng phần mềm SPSS để phân tích phương sai biến đơn.

Bài 6. Có một chiếc túi đựng các hạt nhựa với 6 màu khác nhau (xanh da trời, hồng, xanh lá cây, vàng, da cam, và đỏ). Giả sử từng mẫu xảy ra là như nhau. Vậy một chiếc túi đựng 56 hạt nhựa hẳn chứa chừng khoảng 9,3 hạt nhựa cho mỗi loại màu.

Màu	Xanh da trời	Hồng	Xanh nước biển	Vàng	Da cam	Đỏ
Số hạt nhựa	3	17	6	10	6	14

Dùng SPSS để kiểm định khi bình phương cho giả thiết vừa nêu trên.

Bài 7. Cho bảng số liệu về chi phí quảng cáo (ad) và doanh số bán hàng (sales):

Ad	150	250	350	450	550	650	750
Sales	750	780	1835	2900	3750	4100	4420

Dùng SPSS để xác định mối quan hệ giữa 2 biến số này là cao hay thấp (tìm hệ số tương quan tự động)

Bài 8. Công ty A có doanh số hàng tháng năm 2007 như sau:

t	1	2	3	4	5	6	7	8	9	10	11	12
Sales	210	224	218	235	250	245	243	253	254	270	260	

Dùng SPSS để tìm hệ số tự tương quan.

Bài 9. Công ty A có doanh số (sales) và giá bán (u_price) trong 10 tháng như sau:

Month	1	2	3	4	5	6	7	8	9	10
Sales	102	62	53	121	101	153	52	121	172	203
u_price	129	203	175	152	163	120	164	145	101	109

Dùng SPSS để phân tích mối quan hệ giữa các biến bằng đồ thị và phân tích hồi quy.

Bài 10. Có nguồn số liệu sau:

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Doanhso	162	221	313	291	285	341	91	371	382	361	3 73	354	411	382
Cpqcao	451	552	823	741	854	971	400	1225	1473	1554	1112	1053	1773	1301

Dùng phần mềm SPSS để:

- 1) Phân tích ma trận tương quan
- 2) Phân tích hồi quy dãy số thời gian

TÀI LIỆU THAM KHẢO

- 1- Chu Văn Mẫn (2001). Ứng dụng tin học trong sinh học. Nhà xuất bản Đại học quốc gia Hà Nội.
- 2- Hoàng Trọng (1997). Ứng dụng SPSS for windows để xử lý và phân tích dữ liệu nghiên cứu. Nhà xuất bản Khoa học kỹ thuật.
- 3- Hoàng Trọng (1999). Phân tích dữ liệu đa biến. Nhà xuất bản Thống kê.
- 4- Hoàng Trọng (2002). Xử lý dữ liệu nghiên cứu với SPSS for windows. Nhà xuất bản Thống kê.
- 5- Hồ Đăng Phúc (2005). Sử dụng phần mềm SPSS trong phân tích số liệu. Nhà xuất bản Khoa học kỹ thuật.
- 6- Lê Khắc Bộ (1997-1998). Bài giảng tin học ứng dụng.
- 7- Lê Khắc Bộ (1999–2000). EXCEL - Ứng dụng trong quản trị và kế toán
- 8- Lê Khắc Bộ (2001). Tin học ứng dụng trong kinh tế.
- 9- Lê Khắc Bộ (2004). Tin học kinh tế.
- 10- Lê Khắc Bộ (2004). SPSS – Phân tích dữ liệu kinh tế - xã hội.
- 11- Lê Khắc Bộ (2005). STATA - Ứng dụng trong xử lý dữ liệu.
- 12- Lê Khắc Bộ (2005). EVIEWS - Ứng dụng phân tích dữ liệu.
- 13- Lê Khắc Bộ (2006). GAMS - Ứng dụng phân tích mô hình kinh tế động.
- 14- Lê Khắc Bộ (2006). LINDO - Giải các bài toán kinh tế
- 15- Lê Khắc Bộ (2006). MINITAB - Ứng dụng phân tích thống kê.
- 16- Lê Khắc Bộ (2007). EXCEL nâng cao cho cao học.
- 17- Lê Khắc Bộ (2007). LIMDEP - Ứng dụng trong kinh tế lượng.
- 18- Lê Khắc Bộ (2007). QM – Phân tích định lượng.
- 19- Lê Ngọc Hương (2006-2007). Bài giảng tin học ứng dụng cho chuyên ngành kinh tế nông nghiệp và kế toán doanh nghiệp.
- 20- Mai Văn Bửu, Nguyễn Phú Hưng, Phạm Vũ Thắng (2001). Tin học trong khoa học quản lý. Nhà xuất bản Khoa học kỹ thuật.
- 21- Ngô Thị Thuận (2000-2007). Bài giảng tin học ứng dụng cho sinh viên đại học, cao học chuyên ngành kinh tế nông nghiệp và quản trị kinh doanh.
- 22- Nguyễn Quang Dong (1999). Phân tích dữ liệu. Tài liệu dùng cho lớp bồi dưỡng giáo viên. Khoa Toán kinh tế. Đại học Kinh tế quốc dân.
- 23- Nguyễn Thị Nhuận. Hệ thống các bài tập cho các chuyên ngành kinh tế nông nghiệp, kế toán doanh nghiệp và quản trị kinh doanh các khóa 40-50.
- 24- Nguyễn Tiến, Nguyễn Văn Hoài, Nguyễn Văn Tâm (1996). Quản lý dữ liệu bằng Excel. Nhà xuất bản Giáo dục.
- 25- Nguyễn Tiến, Nguyễn văn Hoài, Nguyễn Văn Tâm (1996). Phân tích kinh doanh bằng Excel. Nhà xuất bản Giáo dục.
- 26- Ông Văn Thông (2000). Thiết lập bảng tính Excel 2000. Phần nâng cao. Nhà xuất bản Thống kê.
- 27- Tin học ứng dụng. Đại học Kinh tế quốc dân. Nhà xuất bản Thống kê, 1999.
- 28- Tô Thế Nguyên (2004-2007). Bài giảng tin học chuyên ngành kinh tế nông nghiệp, kế toán doanh nghiệp và quản trị doanh nghiệp.
- 29- Trần Dũng Tâm, Phạm Tuyết Mai (2000). Hướng dẫn thực hành các bài tập Excel 2000 trong phân tích tài chính và kỹ thuật. Nhà xuất bản Thống kê.
- 30- Trần Trí Dũng (2005). Excel-solver cho kỹ sư. Nhà xuất bản khoa học và kỹ thuật.

- 31- Trần Văn Minh (2001). Xác suất thống kê và các tính toán trên Excel. Nhà xuất bản Giao thông vận tải.
- 32- Trương Văn Tú (2006). Tính toán kinh tế và quản trị kinh doanh trên Excel. Nhà xuất bản Đại học Kinh tế quốc dân.
- 33- Võ Văn Huy, Huỳnh Ngọc Liễu (1998). Máy tính trong kinh doanh - Giải quyết vấn đề ra quyết định. Nhà xuất bản Khoa học kỹ thuật.
- 34- Bary Render, Ralph M. Stair (1997). Quantitative Analysis for Management. Prentice Hall, Inc.
- 35- Lawrence C. Hamilton (2006). Statistics with STATA. Thomson Books/Cole.

MỤC LỤC

	<i>Trang</i>
Chương I. Nhập môn tin học ứng dụng	1
1. Đối tượng nghiên cứu của môn học	1
1.1 Một số khái niệm	1
1.2 Đối tượng, phạm vi ứng dụng tin học trong kinh tế và quản trị kinh doanh	1
1.3 Phương pháp, kỹ thuật ứng dụng tin học trong kinh tế - quản trị kinh doanh	2
2. Phần mềm ứng dụng trong kinh tế và quản trị kinh doanh	2
2.1 Các phần mềm chung	2
2.2 Phần mềm Excel	2
2.3 Phần mềm SPSS	7
Chương II. Ứng dụng phần mềm Excel nâng cao trong kinh tế và quản trị kinh doanh	25
1. Các hàm ứng dụng trong Excel	25
1.1 Các hàm khấu hao	25
1.2 Các hàm tính toán hiệu quả vốn đầu tư	28
1.3 Một số hàm thông dụng khác	37
2. Các lệnh ứng dụng trong Excel	41
2.1 Cụm lệnh Tools/Data Analysis - Cụm lệnh phân tích số liệu	41
2.2 Các lệnh tổng hợp và xử lý số liệu	62
2.3 Lệnh giải bài toán quy hoạch tuyến tính (Tools/Solver)	70
3. Ứng dụng một số chức năng của Excel trong quản trị kinh doanh	77
3.1 Lệnh Tools/Goal Seek	77
3.2 Lệnh Data/Table	78
3.3 Dùng lệnh Tools/Solver để lập kế hoạch sản xuất, phân tích sản xuất	80
Bài tập thực hành chương II	84
Chương III. Ứng dụng SPSS phân tích dữ liệu trong kinh tế và quản trị kinh doanh	92
1. Khảo sát phân phối của một biến và mối quan hệ giữa các biến	92
1.1 Biến định tính	92
1.2 Biến định lượng	93
1.3 Mối quan hệ giữa biến định tính và biến định lượng	95
2. Phân tích phương sai	99
2.1 Phân tích phương sai đa nhân tố lên một biến phụ thuộc	99
2.2 Phân tích phương sai đa nhân tố lên nhiều biến phụ thuộc	101
3. Phân tích hồi quy	103
3.1 Hồi quy tuyến tính đơn (Simple Linear Regression)	103
3.2 Hồi quy tuyến tính bội (Multiple Linear Regression)	106
Bài tập thực hành chương III	111
Tài liệu tham khảo	114