HƯỚNG DẪN SỬ DỤNG SPSS ỨNG DỤNG TRONG NGHIÊN CỨU MARKETING

Lê Văn Huy, Ph.D Candidate (2007) – Danang University of Economics, Email: *huy.lv@due.edu.vn*

ỨNG DỤNG TIN HỌC VÀO PHÂN TÍCH DỮ LIỆU TRONG NGHIÊN CỨU MARKETING

Ngày nay, việc ứng dụng tin học để phân tích dữ liệu trong nghiên cứu marketing là hết sức phổ biến. Có một số phần mềm được sử dụng để phân tích dữ liệu trong nghiên nghiến marketing, mỗi loại đều có những ưu nhược điểm nhất định. Do vậy, cần xác định phần mềm nào được sử dụng trong quá trình phân tích để đạt được hiệu quả cao nhất.

Trong khuôn khổ học phần này, chúng tôi sẽ giới thiệu phần mềm SPSS FOR WINDOWS (Statistical Package for Social Sciences) để phân tích dữ liệu. Ưu điểm của phân mềm này là tính đa năng và mềm dẻo trong việc lập các bảng phân tích, sử dụng các mô hình phân tích đồng thời loại bỏ một số công đoạn (bước) không cần thiết mà một số phân mềm khác gặp phải.
Để đạt được kết quả như mong muốn, cần phải:

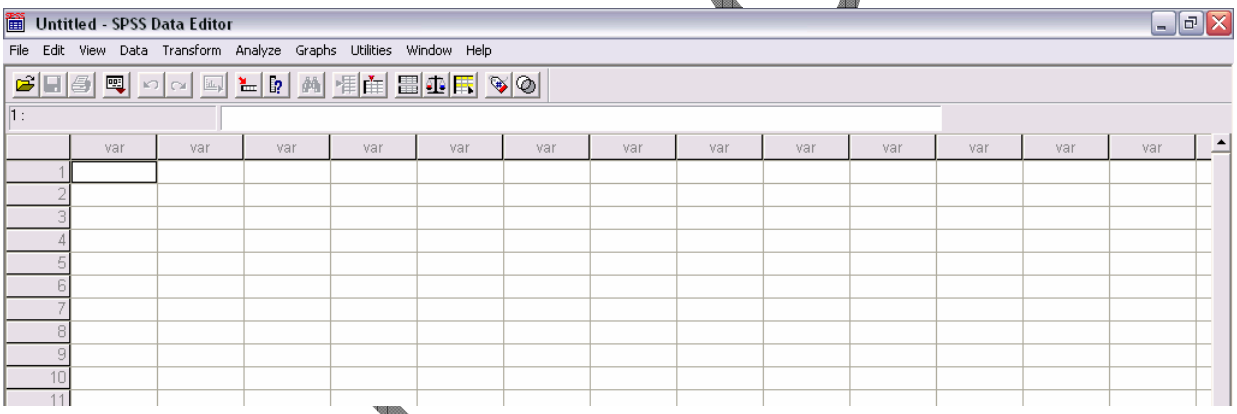
- Nắm vững mục tiêu nghiên cứu dự án
- Nắm vững và tuân thủ những cam kết của dự án về thời gian, chi phí, nguồn nhân lực...

Trên cơ sở xác định bảng câu hỏi và mô hình phân tích (kế hoạch phân tích dữ liệu), quá trình nhập liệu và phân tích có thể thông qua một số công đoạn như sau:

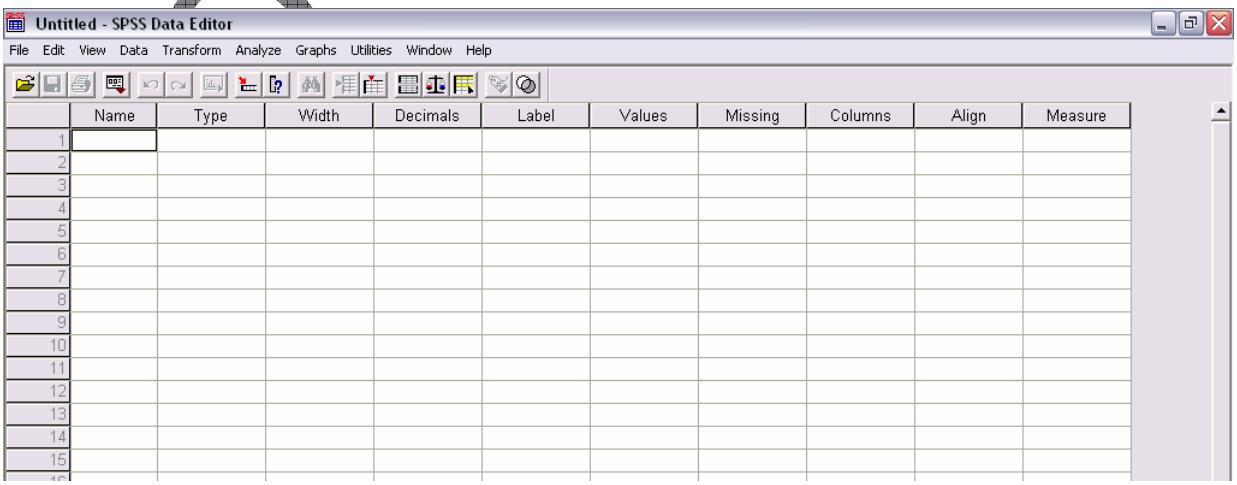
NHẬP LIỆU:

Giao diện nhập liệu

Kích hoạt SPSS, chúng ta thấy giao diện của SPSS như sau:



hoặc:



Lê Văn Huy, Ph.D Candidate (2007) – Danang University of Economics, Email: *huy.lv@due.edu.vn*

Trong đó:

+ **Variable Name**: tên biến (dài 8 kí tự và không có kí tự đặc biệt)

+ **Type**: kiểu của bộ mã hóa

+ **Labels**: nhãn của biến, trong phần này chúng ta có thể nhập nhiều giá trị của nhãn phù hợp với thiết kế của bảng câu hỏi. Sau khi nhập xong mỗi trị của mã hoá, nhấn Add để lưu lại các giá trị trên.

+ **Value**: Giá trị của từng giá trị mã hóa (value) tương ứng với nhãn giá trị (value label) của nó.

+ **Missing**: ký hiệu câu trả lời đúng ra phải trả lời nhưng bị bỏ qua (lỗi), chú ý là giá trị này phải có nét đặc thù riêng biệt so với giá trị khác để dễ dàng phân biệt trong quá trình tính toán.

+ **Column**: thiết đặt độ lớn của cột mang tên biến và vị trí nhập liệu của biến này.

+**Measure**: thang đo lường. Trên cơ sở 4 cấp độ thang đo lường (biểu danh, thứ tự, khoảng cách và tỉ lệ), SPSS sẽ phân ra thành 3 thang đo (biểu danh (nominal), thứ tự (ordinal) và scale (khoảng cách và tỉ lệ).

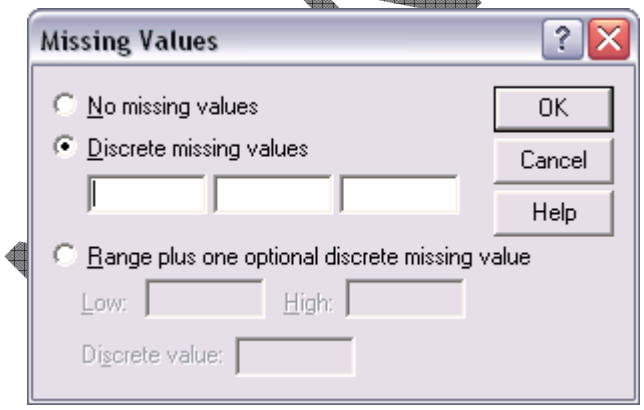
Một số chú ý khi nhập liệu

Nhập giá trị khuyết

Trong quá trình phỏng vấn, có những câu hỏi mà đúng ra được được phỏng vấn phải trả lời câu hỏi đó, tuy nhiên, do một số nguyên nhân, người được phỏng vấn bỏ qua một hoặc vài câu hỏi (hoặc câu trả lời) gọi là giá trị khuyết.

Để đảm bảo thông tin trong quá trình phân tích, chúng ta cần phải định nghĩa những giá trị này như sau: Nhấn **Missing** - Hộp hội thoại **Missing Values** xuất hiện.

- Nhấn **Discrecte missing values**, đặt các trị missing values vào các ô trống, trị được nhập tại các ô trống sẽ đại diện cho những giá trị khuyết.
- Chúng ta có thể định nghĩa các giá trị khuyết theo một khoảng giá trị nào đó bằng các nhấn và nhập liệu vào **Range plus one optional discrete missing value**.
- Tất cả các giá trị khuyết sẽ không tham gia vào quá trình phân tích.



Chèn một biến mới hoặc bảng ghi mới

- Nhấn **Data/Insert Variable**
- Nhấn **Data/Insert Case**
- Tìm đến bảng ghi cần thiết: **Go to Case**

Lê Văn Huy, Ph.D Candidate (2007) – Danang University of Economics, Email: *huy.lv@due.edu.vn*

Sắp xếp bảng ghi

- Nhấn **Sort Case**
- Sắp xếp theo biến tại **Sort by** với chiều tăng (Ascending) hoặc giảm (Descending)

Biến một biến thành một bảng ghi

- Nhấn **Data/Transpose**
- Variable(s) là những biến cần thay đổi

Kiểm tra giá trị nhập

- Nhãn toàn bộ giá trị: **Nhấn View/ Value Labels**
- Kiểm tra một biến nào đó: **Utilities/Variables**
- Kiểm tra bộ mã hoá **Utilities/File Info**, với bộ mã hoá này, ta có thể kiểm tra lại một lần nữa công việc định nghĩa các biến hoặc cũng có thể làm danh bạ cho việc nhập số liệu sau này.

Tạo biến mới không hoặc có ràng buộc một điều kiện

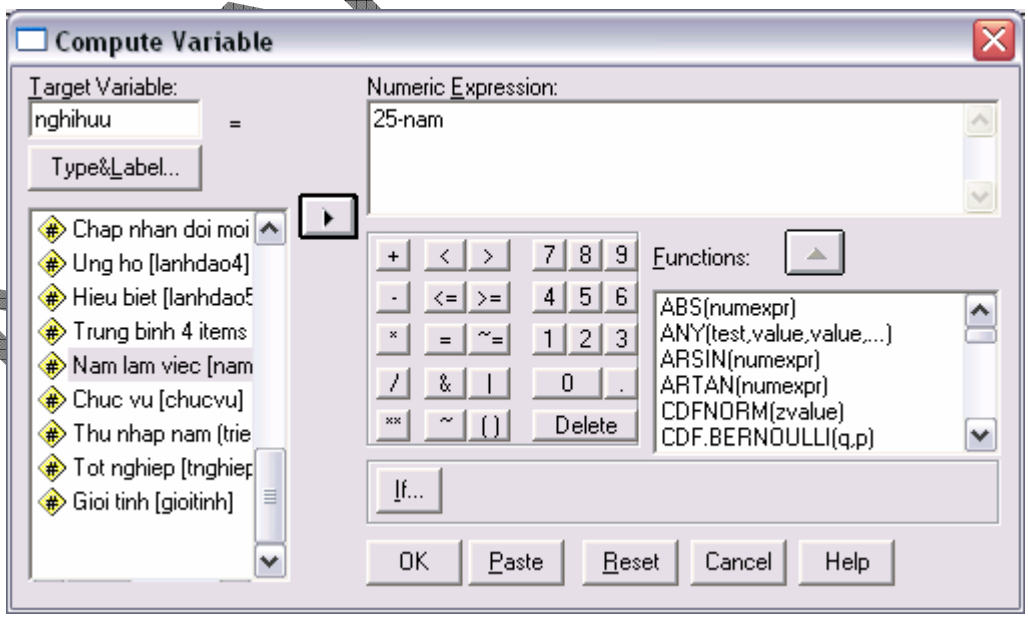
Trong quá trình nhập liệu, để có thể rút ngắn thời gian nhập liệu hoặc để phục vụ mục đích phân tích, chúng ta còn có thể tạo ra biến mới từ các dữ kiện và cấu trúc của biến đã nhập.

- Tạo biến mới không điều kiện: Giả sử theo số liệu thống kê như trên, để biết được số năm công tác còn lại trước khi nghỉ hưu là bao nhiêu năm nữa (giả sử mỗi lao động được nghỉ hưu sau 25 năm công tác). Như vậy ta thành lập một biến mới *nghihuu* sẽ bằng 25-nam
  + Nhấn **Transform/Compute**

  + Trong ô **Target Variable** nhập biến mới (*nghihuu*), trong đó chúng ta cần phải định nghĩa **Type&Label** để tiện cho việc quản lí và so sánh các giá trị sau này.

  + Trong ô **Numeric Expression** nhập giá trị cần gán cho biến mới từ biến đích cho trước.

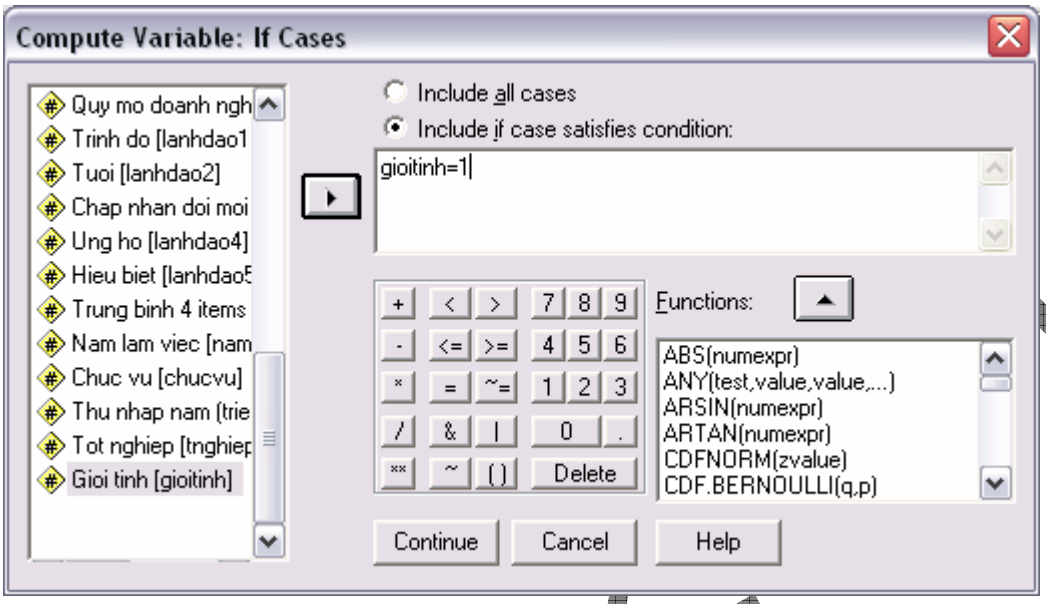
  Chú ý: Khi gặp các biến thuộc kiểu chuỗi, ngày tháng... chúng ta cần phải tìm một hàm tương ứng để quy các giá trị này về giá trị tương đồng mà chúng ta có thể so sánh được (sử dụng hàm Function)



- Tạo biến mới có điều kiện: Cũng như ví dụ trên nhưng chúng ta cần phân chia ra thành nam và nữ thì sau khi thiết đặt các giá trị như trên xong.

Lê Văn Huy, Ph.D Candidate (2007) – Danang University of Economics, Email: *huy.lv@due.edu.vn*

- Nhấn **If** tiếp theo nhấn **Include if case satisfies condition** trong hộp hội thoại để thiết đặt điều kiện (áp dụng cho những người có giới tính là nam thì điều kiện thiết đặt là gioitinh=1 như trong hộp hội thoại:



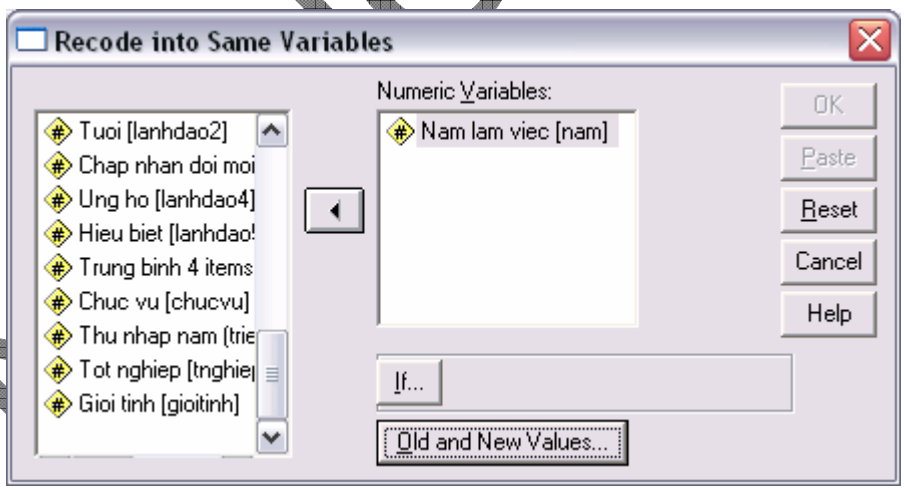
Mã hoá lại biến:

Trong một số trường hợp, do nhu cầu của quá trình phân tích, chúng ta cần phải mã hóa lại các biến. Có hai hình thức mã hoá như sau:

- Mã hoá dùng lại tên biến cũ:

  + Nhấn Transform/Recode/Into Same Variables

  + Đưa biến cần mã hoá lại vào ô Numeric Variable



  + Nhấn **If** để thiết đặt các điều kiện (nếu có)

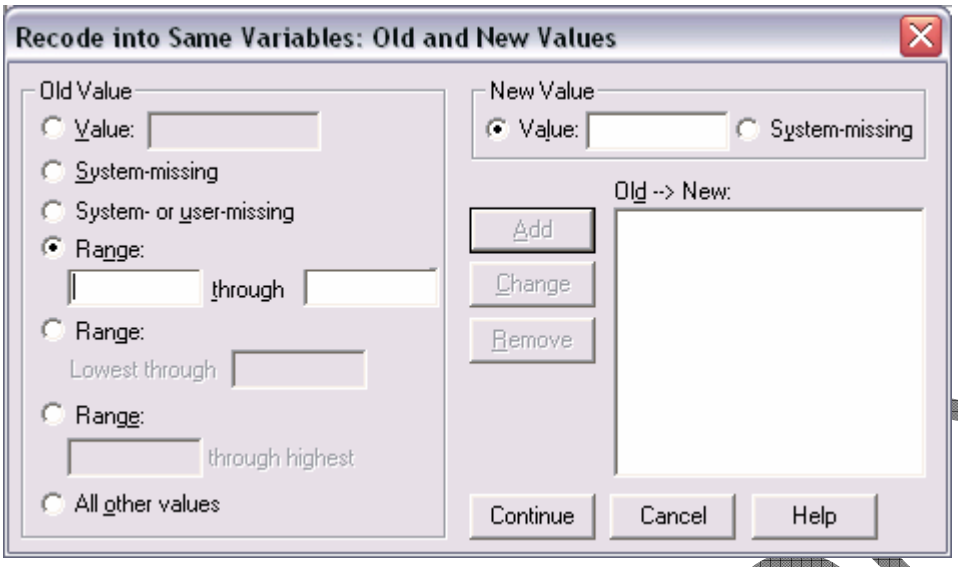
  + Nhấn **Old and New Values** để thay đổi bộ mã hoá

    * Trong ô **Old Value** là giá trị cũ, và **New Value** là giá trị mới cần nhập

    * Nếu nhập giá trị mới ở thang điểm biểu danh, khoảng cách, tỷ lệ thì nhập tại ô Value.

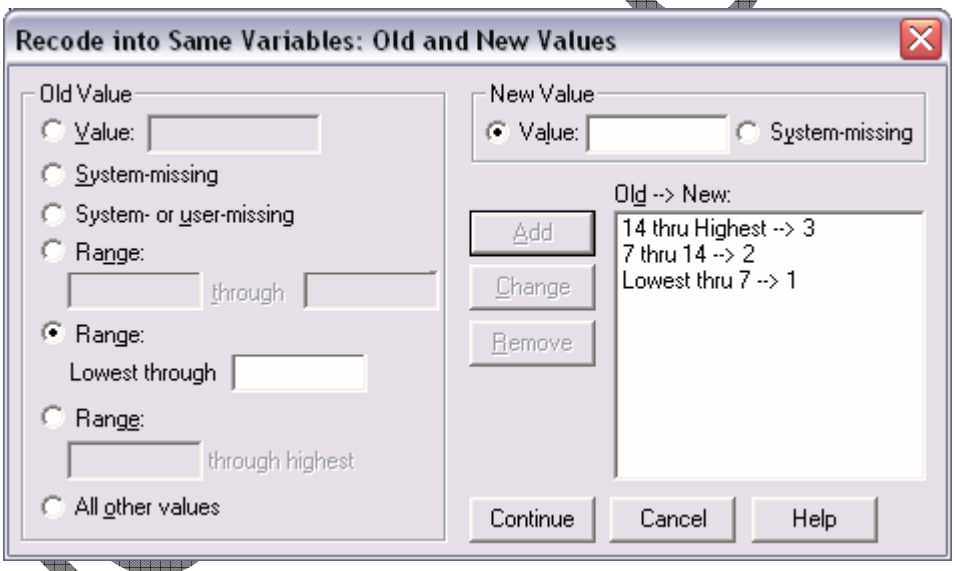
    * Nếu mã hoá giá trị với thang điểm khoảng cách - Nhấn Range

Lê Văn Huy, Ph.D Candidate (2007) – Danang University of Economics, Email: *huy.lv@due.edu.vn*



Ví dụ: Để phục vụ cho việc phân tích, ta mã hoá lại tuổi của sinh viên theo thang điểm khoảng cách như sau:

- *1 : Dưới 7 năm*
- *2 : Từ 7 đến 14 năm*
- *3 : Trên 14 năm*



** Giá trị trên 14 năm bấm Range/throught Highest và nhập liệu*

** Giá trị dưới 7 năm bấm Range/Lowest throught và nhập liệu*

** Có thể giữ nguyên giá trị khuyết hay cần thay đổi, nếu giữ nguyên cần chú ý là giá trị đó có rơi vào các trường hợp mã chúng ta mã hoá không để khỏi ảnh hưởng đến các giá trị phân tích.*

- Mã hoá dùng lại không dùng tên biến cũ (lưu trên biến mới):

+ Nhấn **Transform/Recode/Into Different Variables**

+Tên biến mới được đặt ở ô **Name** với các thông số thoả mãn một biến bình thường.

+ Nhãn của biến được thiết đặt tại ô **Label**, sau đó nhấn **Change** để lưu.

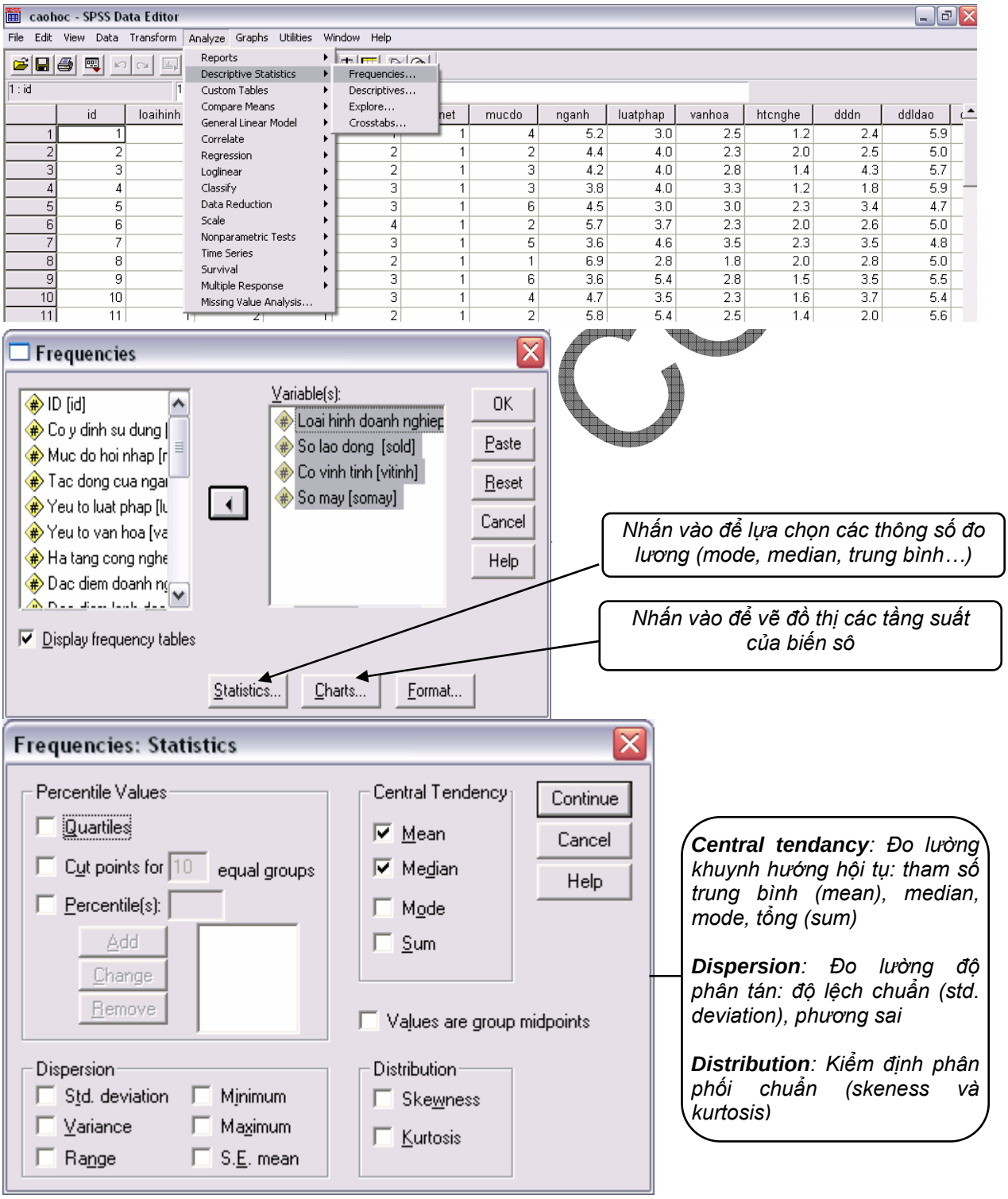
Lê Văn Huy, Ph.D Candidate (2007) – Danang University of Economics, Email: *huy.lv@due.edu.vn*

+ Các thông số khác được thực hiện như ở mã hoá dùng lại biến cũ.

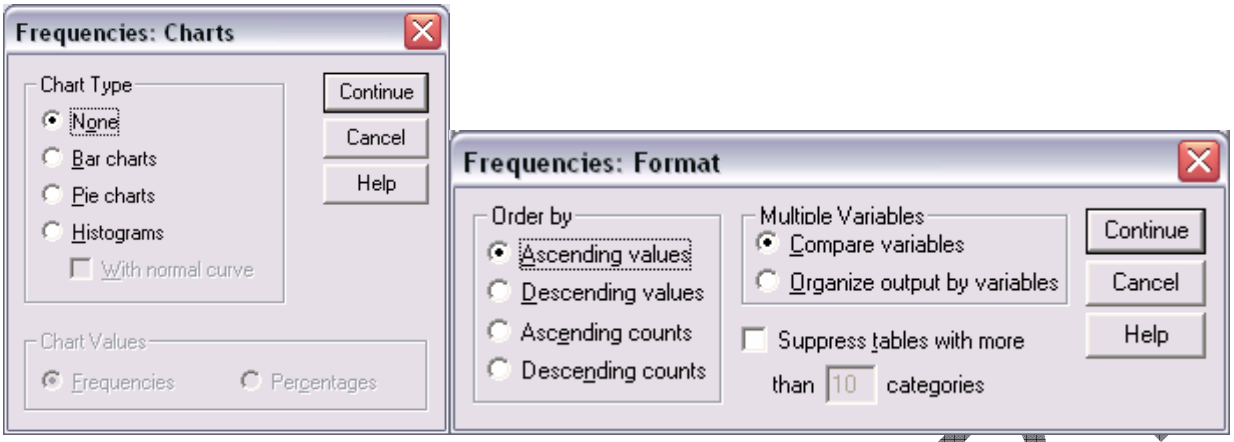
PHÂN TÍCH MÔ TẢ (THỐNG KÊ MÔ TẢ):

Bảng phân bố tần suất

Bảng phân phối tầng suất được thể hiện với tất cả các biến **định tính** (rời rạc) với các thang đo biểu danh, thứ tự và các biến **định lượng** (liên tục) với thang đo khoảng cách hoặc tỉ lệ.



Lê Văn Huy, Ph.D Candidate (2007) – Danang University of Economics, Email: *huy.lv@due.edu.vn*



Tần suất xuất hiện

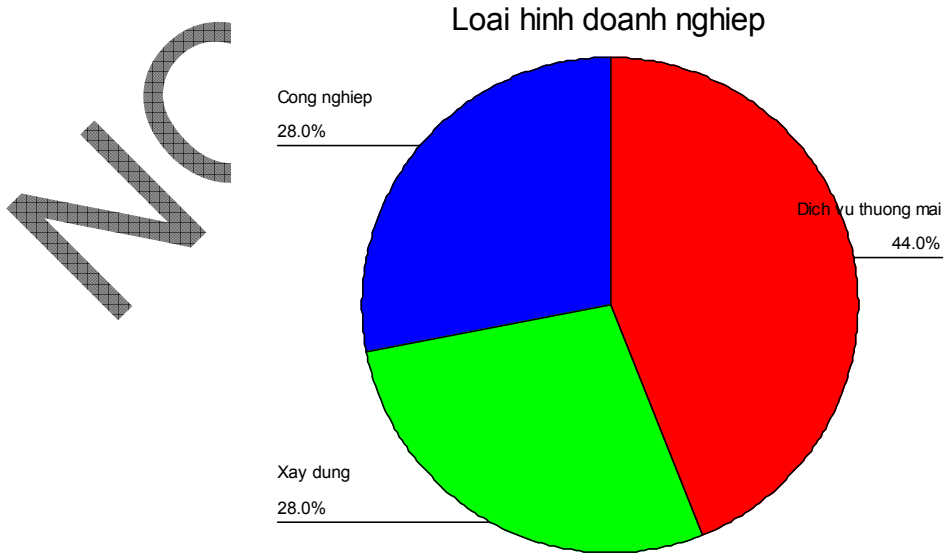
Tỷ lệ phần trăm

Loai hinh doanh nghiep

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Dich vu thuong mai	88	44.0	44.0	44.0
	Xay dung	56	28.0	28.0	72.0
	Cong nghiep	56	28.0	28.0	100.0
	Total	200	100.0	100.0	

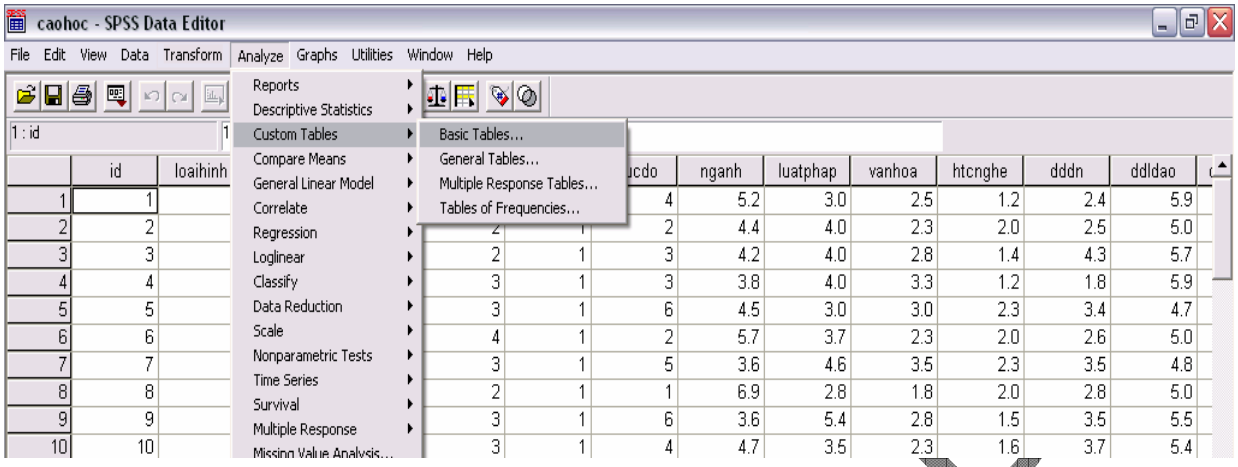
So lao dong

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Tu 1 den 5	25	12.5	12.5	12.5
	Tu 6 den 20	61	30.5	30.5	43.0
	Tu 21 den 200	63	31.5	31.5	74.5
	Tu 200 den 300	45	22.5	22.5	97.0
	Tren 300	6	3.0	3.0	100.0
	Total	200	100.0	100.0	

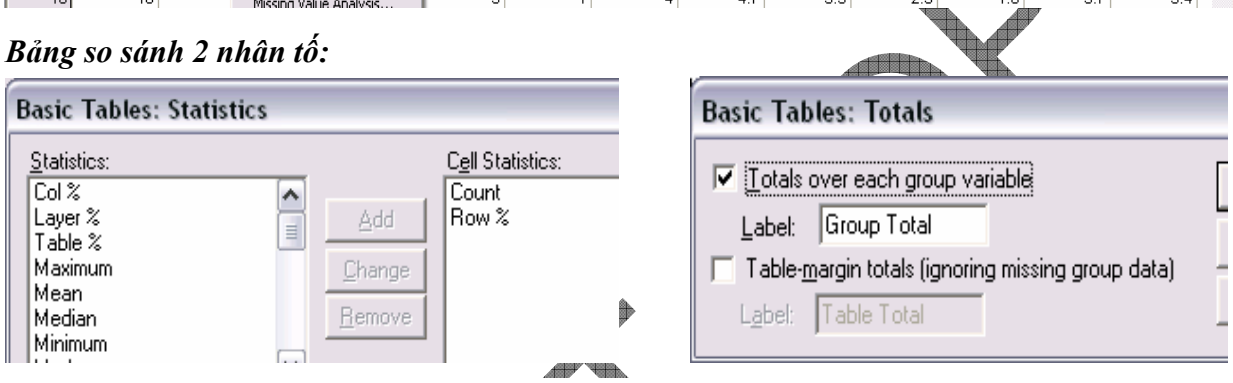


Lê Văn Huy, Ph.D Candidate (2007) – Danang University of Economics, Email: *huy.lv@due.edu.vn*

Lập bảng so sánh



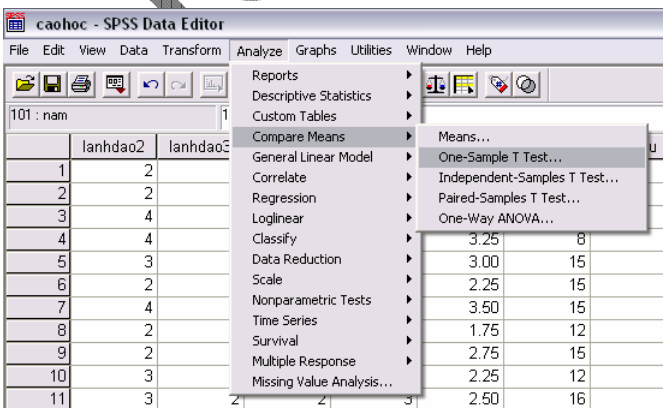
Bảng so sánh 2 nhân tố:

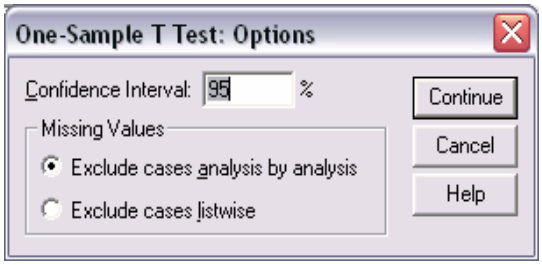


		Loai hinh doanh nghiep					
		Dich vu thuong mai		Xay dung		Cong nghiep	
		Count	Row %	Count	Row %	Count	Row %
So lao dong	Tu 1 den 5	7	28.0%	6	24.0%	12	48.0%
	Tu 6 den 20	26	42.6%	21	34.4%	14	23.0%
	Tu 21 den 200	26	41.3%	19	30.2%	18	28.6%
	Tu 200 den 300	27	60.0%	7	15.6%	11	24.4%
	Tren 300	2	33.3%	3	50.0%	1	16.7%
Group Total		88	44.0%	56	28.0%	56	28.0%

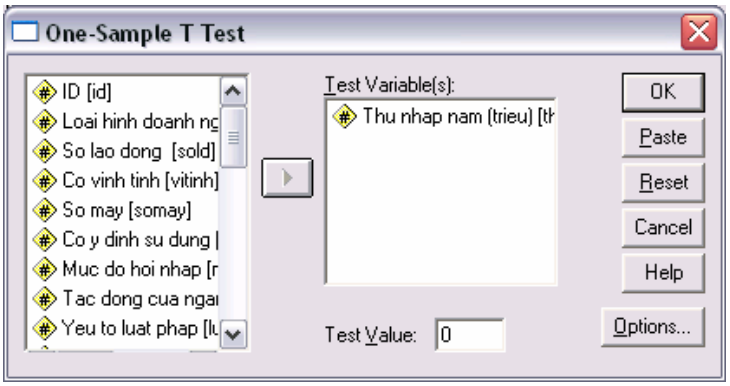
Phân tích một biến định lượng

Ước lượng tham số trung bình (một nhóm)





Lê Văn Huy, Ph.D Candidate (2007) – Danang University of Economics, Email: *huy.lv@due.edu.vn*



Giá trị trung bình

Độ lệch chuẩn

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Thu nhap nam (trieu)	200	33224.00	12932.72	914.48

Giới hạn trên của ước lượng

Giới hạn dưới của ước lượng

One-Sample Test

	Test Value = 0					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference: Lower	Upper
Thu nhap nam (trieu)	36.331	199	.000	33224.00	31420.68	35027.32

Ước lượng sự khác biệt giữa hai tham số trung bình (độc lập hoặc phụ thuộc)

KIỂM ĐỊNH THAM SỐ

Kiểm định t đối với tham số trung bình mẫu

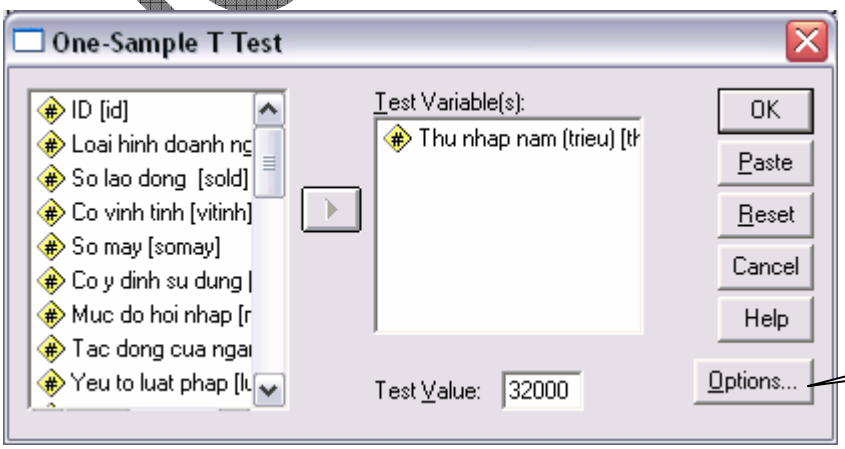
Như chúng ta đã biết, thu nhập trung bình của các đối tượng phỏng vấn là 33,224 triệu/năm, có giả thiết cho rằng thu nhập của đối tượng mà chúng ta phỏng vấn trên tổng thể là 32 triệu/năm, chúng ta cần kết luận nhận định đó có đúng không.

Khi đó, giả thiết của bài toán là:

$H_0 : \mu = \mu_0 = 32$ (triệu) và $H_1: \mu \neq \mu_0 = 32$ (triệu)

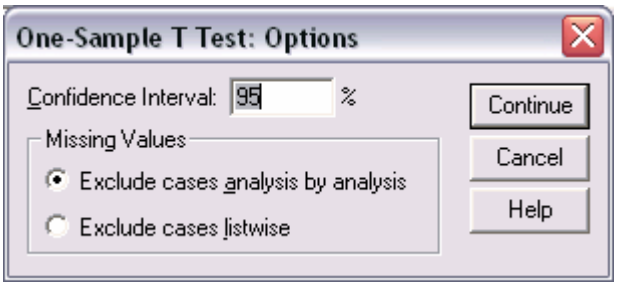
☞ Nhấn **Analyze – Compare Means – One sample T test**.

☞ Chọn biến cần phân tích vào ô **Test Variable(s)**, đặt giá trị μ_0 vào ô **Test Value**.



Nhấn **Option** để thiết đặt độ tin cậy (giả sử đ tin cậy là 95%)

Lê Văn Huy, Ph.D Candidate (2007) – Danang University of Economics, Email: *huy.lv@due.edu.vn*



☞ Bấm Continue và bấm OK ở hộp hội thoại ban đầu, kết quả thu được như sau:

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Thu nhap nam (trieu)	200	10750	82500	33224.00	12932.72
Valid N (listwise)	200				

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Thu nhap nam (trieu)	200	33224.00	12932.72	914.48

One-Sample Test

	Test Value = 32000					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Thu nhap nam (trieu)	1.34	199	.182	1224.00	-579.32	3027.32

Giá trị t-student = 1,34

Giá trị p-value =0,182>0,05

☞ Tại các biểu trên, ta có thể biết giá trị trung bình, độ lệch chuẩn của mẫu. Ngoài ra t=1,34 nên p-value=0,182>0,05 nên chúng ta chưa có cơ sở để bác bỏ H_0 hay chưa có cơ sở để chấp nhận H_1.

Kiểm định tham số trung bình hai mẫu (hai mẫu độc lập)

Giả sử ta muốn so sánh thu nhập trung bình giữa những người có giới tính nam và nữ trên tổng thể có khác nhau hay không, ta có giả thiết:

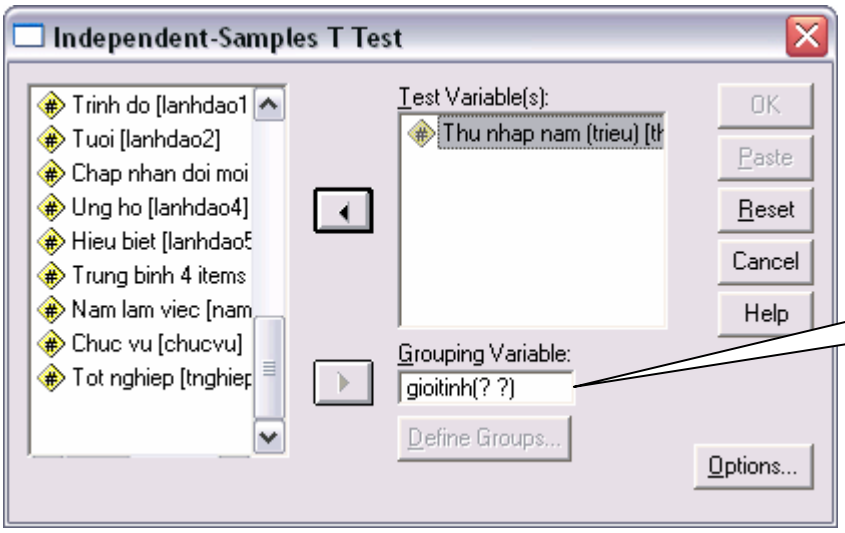
H_0: Thu nhập trung bình của người nam và người nữ bằng nhau trên tổng thể

H_1: Thu nhập trung bình của người nam và người nữ không bằng nhau trên tổng thể

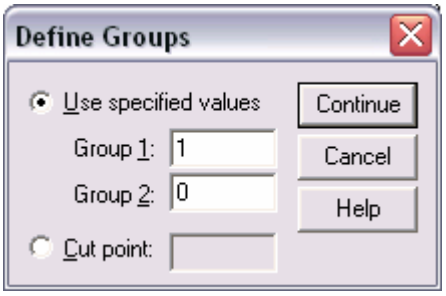
☞ Nhấn **Analyze – Compare Means – Independent sample t-test**.

☞ Chọn biến thunhap vào ô **Test Variables** và biến gioitinh vào ô **Grouping Variable**

Lê Văn Huy, Ph.D Candidate (2007) – Danang University of Economics, Email: *huy.lv@due.edu.vn*



Nhấn vào **Define Groups** để định nghĩa các nhóm với Nam=1 và Nữ = 0



Nhấn vào **Define Groups** để định nghĩa các nhóm với Nam=1 và Nữ = 0

☞ Kết quả như sau

Group Statistics

	Gioi tinh	N	Mean	Std. Deviation	Std. Error Mean
Thu nhap nam (trieu)	Nam	124	37053.23	13962.42	1253.86
	Nu	76	26976.32	7763.42	890.52

Trung bình người có giới tính là Nữ

Trung bình người có giới tính là Nam

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Thu nhap nam (trieu)	Equal variances assumed	17	.000	5.77	198	.000	10076.91	1747.75	6630	13524
	Equal variances not assumed			6.55	196.4	.000	10076.91	1537.92	7044	13110

Nếu sig. trong kiểm định phương sai<0,05 thì phương sai giữa hai mẫu không bằng nhau, ta sẽ dùng kết quả kiểm định t ở dòng thứ 2

Giá trị t của kiểm định

p-value của giá trị t

☞ Kiểm định Leneve's (giả thiết H_0: phương sai của hai mẫu (biến) bằng nhau, H_1: phương sai của hai mẫu (biến) không bằng nhau) sẽ cho phép kiểm định phương sai hai mẫu có bằng nhau

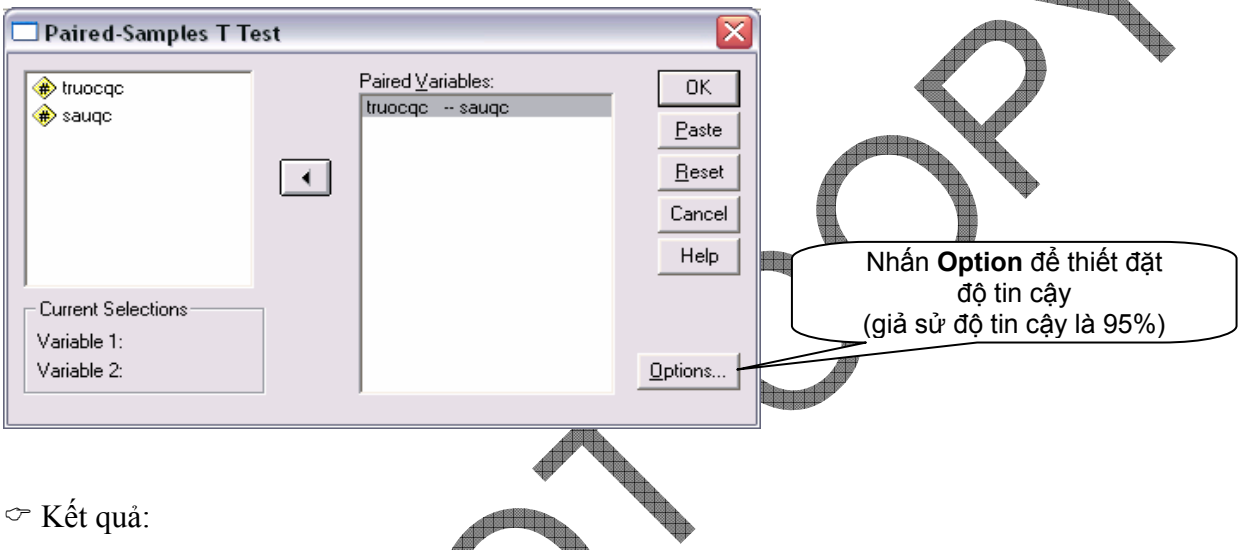
Lê Văn Huy, Ph.D Candidate (2007) – Danang University of Economics, Email: *huy.lv@due.edu.vn*

hay không, trong trường hợp này nếu sig. của F (trong thống kê Leneve's) < 0,05 ta bác bỏ H_0, chấp nhận H_1 nghĩa là phương sai của hai mẫu không bằng nhau, do vậy giá trị t mà ta phải tham chiếu là giá trị t ở dòng thứ 2. Ngược lại nếu sig. >0,05 thì phương sai của hai mẫu bằng nhau, ta sẽ dùng kết quả kiểm định t ở dòng thứ nhất.

☞ Đối với kiểm định t, ta nhận thấy rằng t=6,55 và p-value = 0,000<0,05 năm ta có thể bác bỏ H_0 và chấp nhận H_1, có nghĩa là thu nhập trung bình giữa người nam và nữ sẽ khác nhau.

Kiểm định tham số trung bình hai mẫu (hai mẫu phụ thuộc)

☞ Nhấn **Analyze – Compare Means – Paired sample t-test**. Chọn biến cần phân tích vào ô **Paired Variables**.



☞ Kết quả:

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	TRUOCQC	42.9333	15	30.6419	7.9117
	SAUQC	44.1333	15	28.1422	7.2663

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	TRUOCQC - SAUQC	-1.200	5.7842	1.4935	-4.4032	2.0032	-.803	14	.435

Giá trị ước lượng (giới hạn dưới)

Giá trị ước lượng (giới hạn trên)

Giá trị t-student = -0,803

Giá trị p-value =0,435>0,05

☞ Vì giá trị t=-0,803 và p-value = 0,435>0,05 nên chúng ta chưa có cơ sở để bác bỏ H_0 tức là chưa có cơ sở để chấp nhận H_1.

Lê Văn Huy, Ph.D Candidate (2007) – Danang University of Economics, Email: *huy.lv@due.edu.vn*

Phân tích phương sai (Analysis of variance – ANOVA)

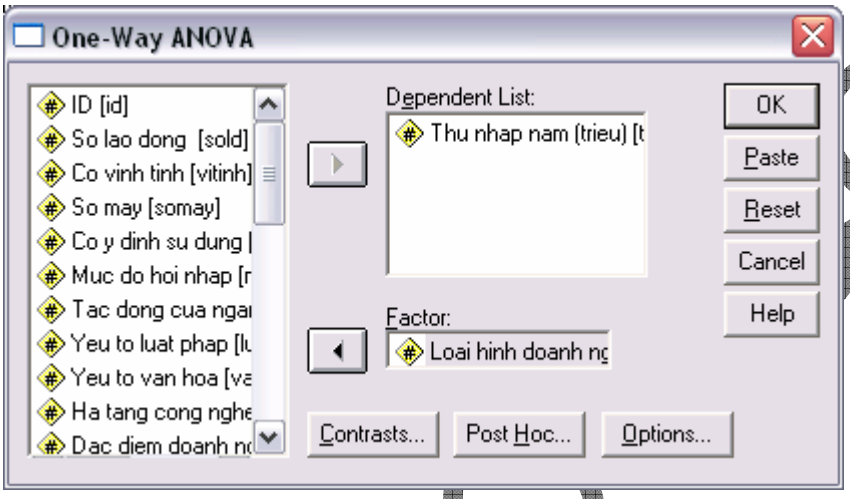
Giả sử chúng ta muốn so sánh thu nhập trung bình của các đối tượng làm trong những lĩnh vực dịch vụ - thương mại, xây dựng và công nghiệp có khác nhau hay không. Giả thiết và đối thiết sẽ là:

H_0: Thu nhập trung bình của những người làm trong lĩnh vực dịch vụ - thương mại, xây dựng và công nghiệp bằng nhau

H_1: Thu nhập trung bình của người làm trong lĩnh vực dịch vụ - thương mại, xây dựng và công nghiệp không bằng nhau (có nghĩa là tồn tại ít nhất một thu nhập trung bình của một ngành khác với ít nhất một thu nhập trung bình của hai ngành còn lại)

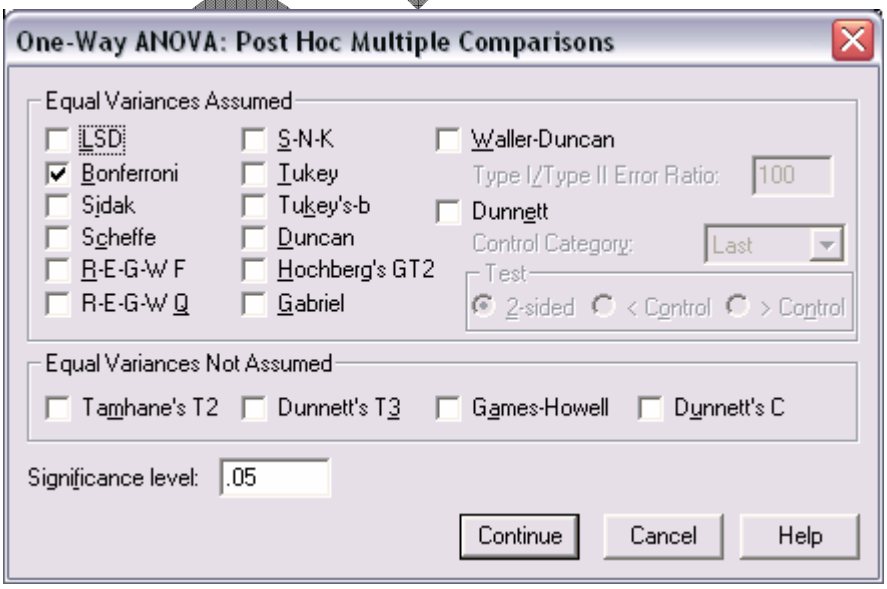
☞ Nhấn **Analyze – Compare Means – One-way ANOVA**.

☞ Chọn biến cần phân tích (định lượng) vào ô **Dependent List** và biến phân loại vào ô **Factor**



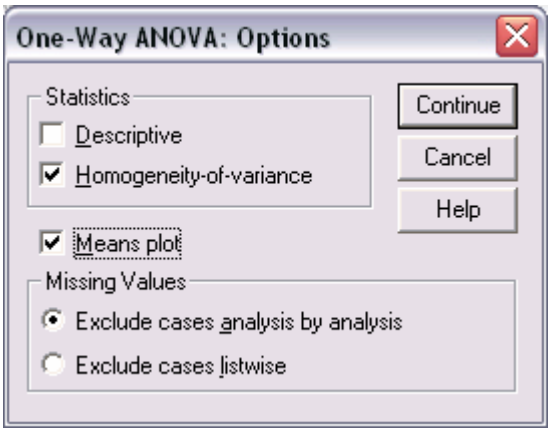
☞ Nhấn **Post Hoc** để chọn loại kiểm định nhằm xác định cụ thể sự khác biệt giữa các nhóm (nhóm nào khác với nhóm nào). Chúng ta có thể chọn Bonferroni hoặc Tukey's-b (hai thống kê này đều cho ra cùng một kết quả).

☞ Nếu phương sai giữa các nhóm cần so sánh không bằng nhau, chúng ta chọn Tamhane's T2 (ứng dụng cho kiểm định t từng cặp nếu phương sai của chúng không bằng nhau).



Lê Văn Huy, Ph.D Candidate (2007) – Danang University of Economics, Email: *huy.lv@due.edu.vn*

☞ Nhấn **Continue**, nhấn **Option** để thiết đặt các lựa chọn.



☞ Trong đó **Homogeneity-of-variance** để kiểm định sự bằng nhau phương sai các nhóm, **Means plot** để làm cho hình minh họa.

Test of Homogeneity of Variances

Thu nhap nam (trieu)

Levene Statistic	df1	df2	Sig.
.414	2	197	.661

☞ Vì Sig. >0,05 nên ta có thể khẳng định là phương sai của các nhóm là bằng nhau, thỏa mãn điều kiện của phân tích ANOVA.

ANOVA

Thu nhap nam (trieu)

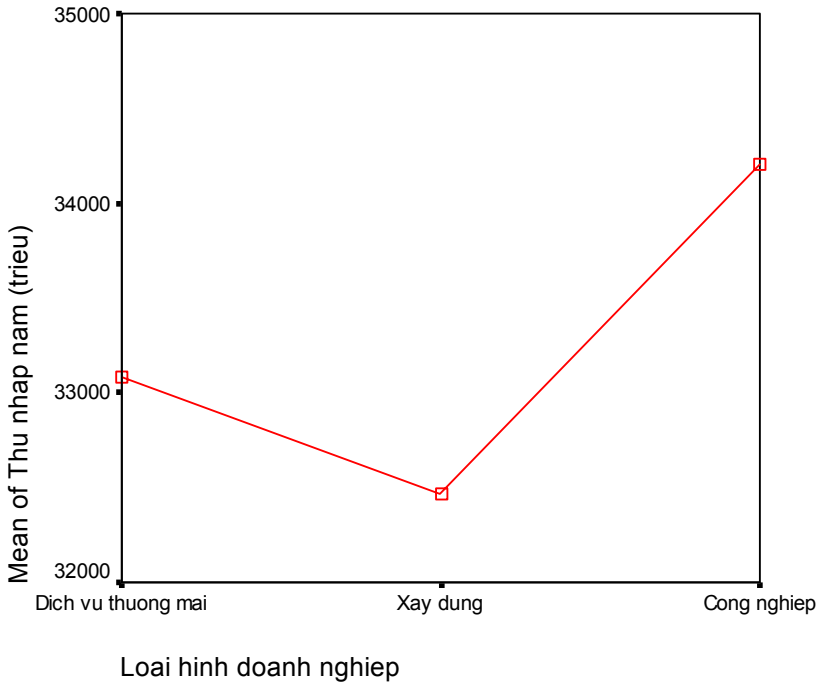
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	87185676.623	2	43592838.312	.259	.772
Within Groups	33196619123.377	197	168510756.971		
Total	33283804800.000	199			

☞ Với F=0,259 và p-value = 0,772>0,05 nên chưa có cơ sở để bác bỏ H_0 hay chưa có cơ sở để chấp nhập H_1

☞ Trong các trường hợp khác, nếu ta bác bỏ H_0 và chấp nhận H_1, với thống kê Bonferonni ta có thể biết được sự khác nhau từng cặp của các tham số trung bình.

☞ Means plots

Lê Văn Huy, Ph.D Candidate (2007) – Danang University of Economics, Email: *huy.lv@due.edu.vn*

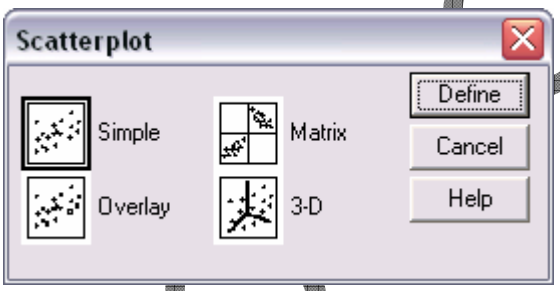


Hồi quy tuyến tính

Giả sử chúng ta mong muốn tìm mối tương quan giữa hai biến năm làm việc (biến độc lập) và thu nhập hàng năm (biến phụ thuộc) trên tổng thể, chúng ta sẽ thực hiện như thế nào.

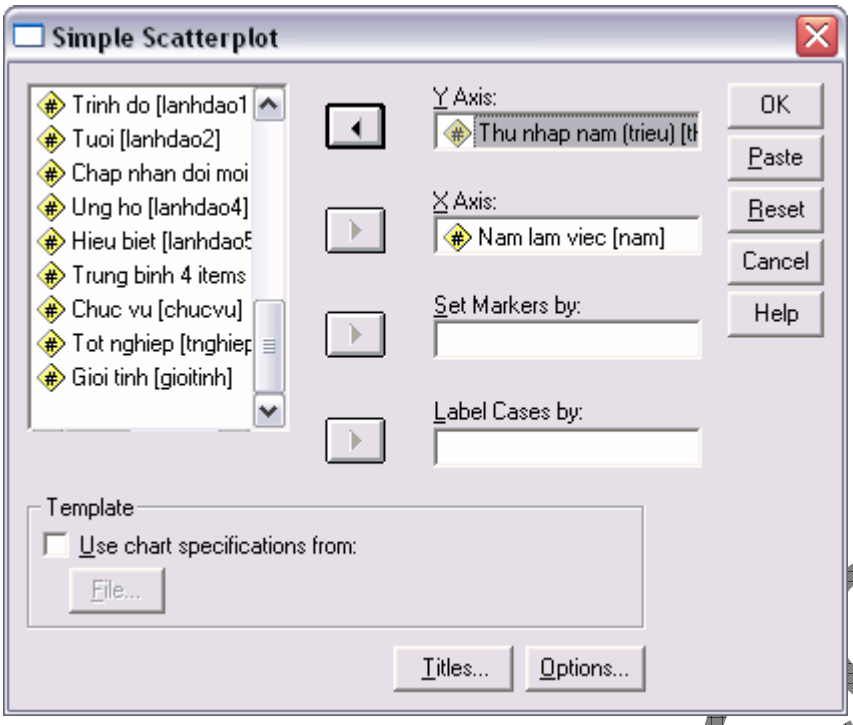
☞ Vẽ sơ đồ, kiểm tra bằng thị giác mối quan hệ

☞ Vào **Graphs**, nhấn **Scatter**

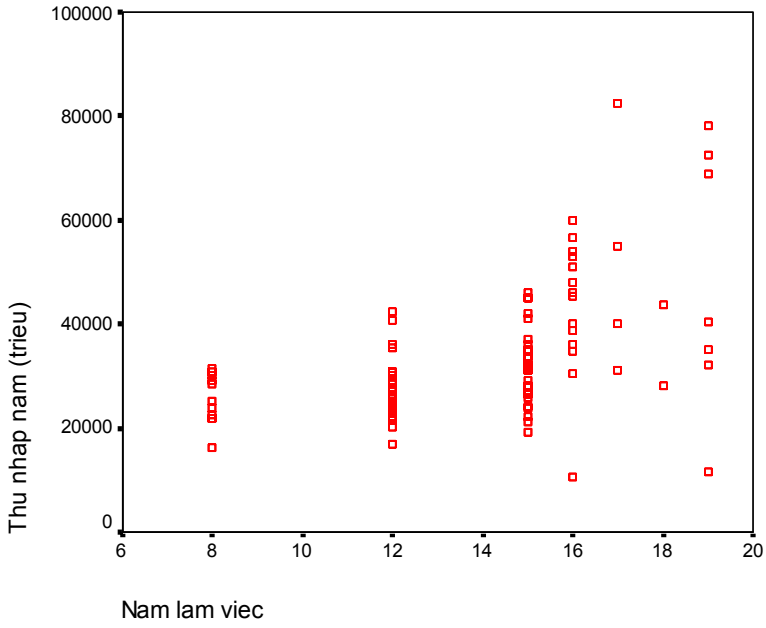


☞ Chọn **Simple** và bấm **Define**

Lê Văn Huy, Ph.D Candidate (2007) – Danang University of Economics, Email: *huy.lv@due.edu.vn*



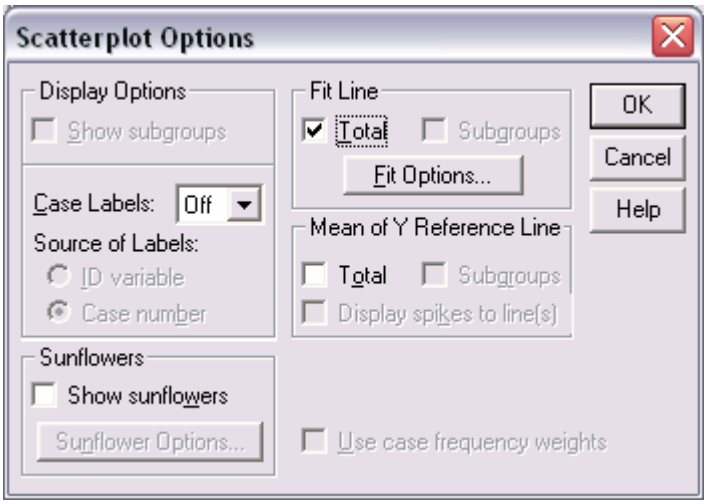
☞ Chọn các biến vào ô **Y Axis** (biến phụ thuộc) và **X Axis** (biến độc lập), bấm **OK**



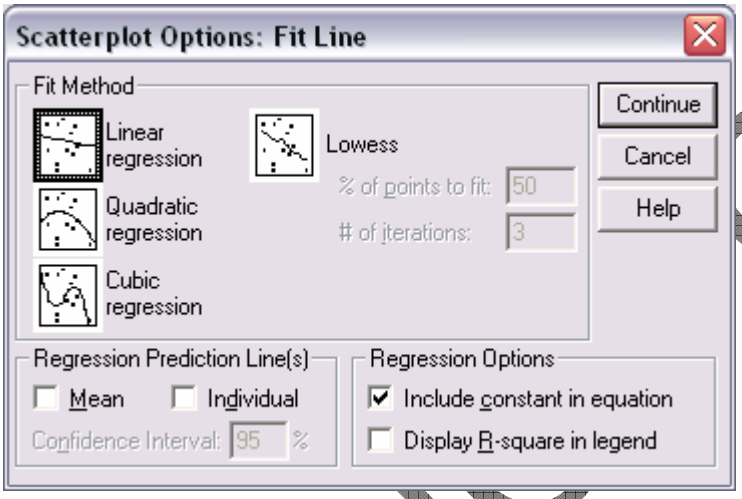
☞ Chúng ta có thể xem đường hồi quy lí thuyết của dãy dữ liệu bằng cách click hai lần vào chuôt.

☞ Sau khi một màn hình mới hiện ra, vào Chart – Option, hội hội thoại tiếp theo sẽ hiện ra – Bấm OK – Hội hội thoại sẽ là:

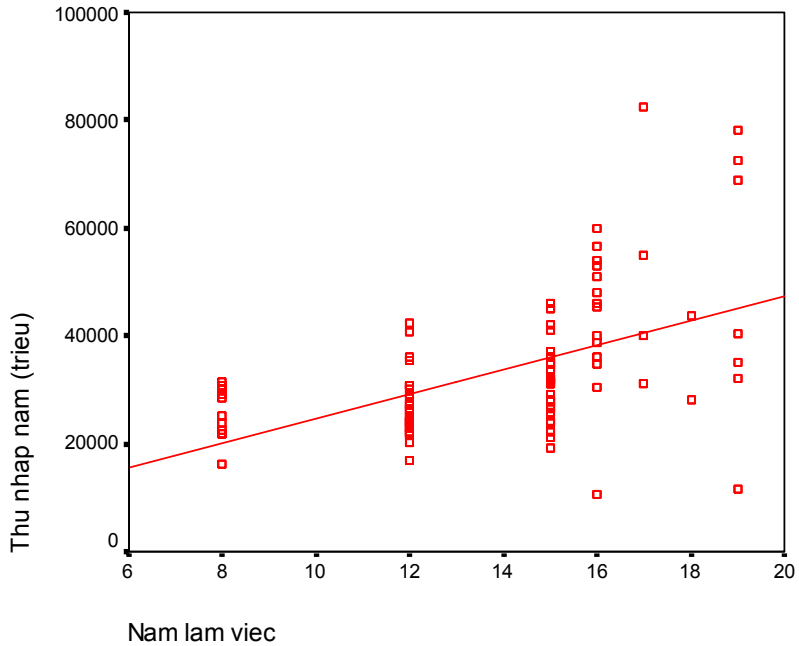
Lê Văn Huy, Ph.D Candidate (2007) – Danang University of Economics, Email: *huy.lv@due.edu.vn*



☞ Bấm **Fit Options** chọn **Linear regression**



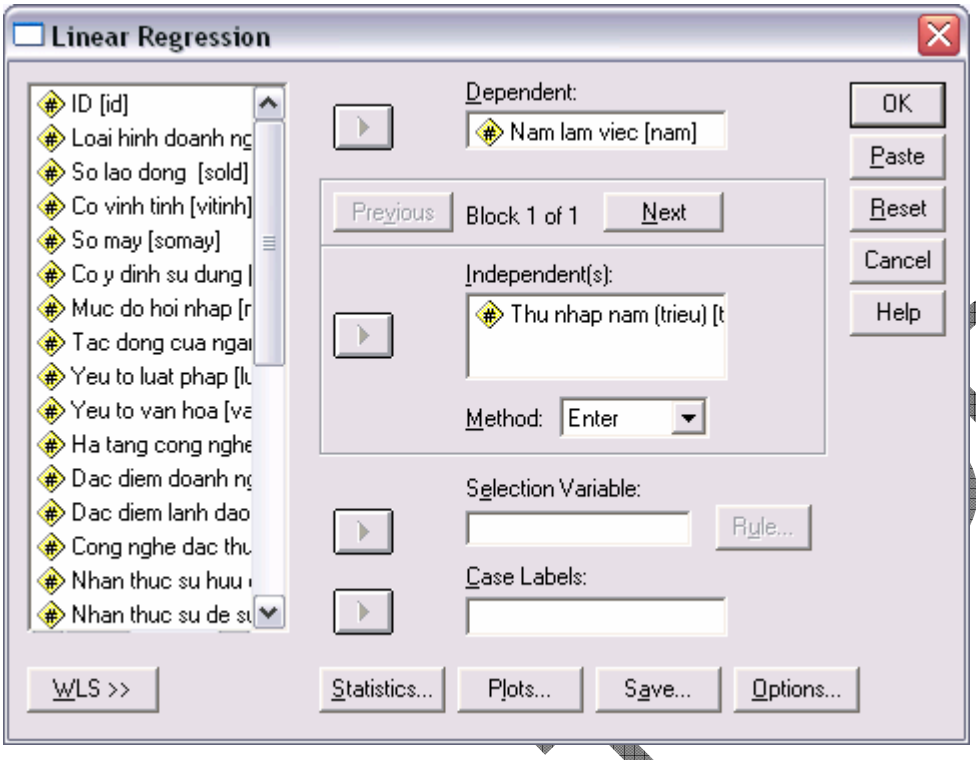
☞ Bấm **Continue** và **OK**



Lê Văn Huy, Ph.D Candidate (2007) – Danang University of Economics, Email: *huy.lv@due.edu.vn*

Rõ ràng trên hình vẽ bên, ta có thể hình dung có mối quan hệ tuyến tính (theo đường thẳng) giữa số năm làm việc và thu nhập/năm. Để kiểm tra một cách chính xác, ta thực hiện thao tác hồi quy.

☞ Vào **Analyze** và **Regression** chọn các biến vào các ô tương ứng



ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	449.294	1	449.294	71.115	.000[a]
	Residual	1250.926	198	6.318		
	Total	1700.220	199			

a. Predictors: (Constant), Thu nhap nam (trieu)

b. Dependent Variable: Nam lam viec

Vì F=71,115 và p-value=0,000 nên chúng ta có thể khẳng định tồn tạo mô hình hay tồn tại mối quan hệ giữa hai biến năm làm việc và thu nhập trên tổng thể.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.514[a]	.264	.261	2.51

a. Predict (Constant), Thu nhap m (trieu)

Hệ số tương quan R

Hệ số xác định R^2

Ta có:

- Hệ số tương quan R đo lường mức độ tương quan giữa hai biến
- Hệ số xác định R^2 đánh giá mức độ phù hợp của mô hình thể hiện mối quan hệ tương quan tuyến tính

R^2 = 0,264 có nghĩa là biến số năm làm việc sẽ giải thích 26,4% thu nhập/ năm của nhân viên (còn lại là những biến số khác).

Lê Văn Huy, Ph.D Candidate (2007) – Danang University of Economics, Email: *huy.lv@due.edu.vn*

Ta có R^2_a=0,261, ta có thể kết luận mối quah hệ giữa hai biến này rất yếu vì R^2_a=0,261<0,3.

- Nếu	R <0,3	- Nếu	R^2 <0,1	Tương quan ở mức thấp
- Nếu	0,3 ≤ R <0,5	- Nếu	0,1 ≤ R^2 <0,25	Tương quan ở mức trung bình
- Nếu	0,5 ≤ R <0,7	- Nếu	0,25 ≤ R^2 <0,5	Tương quan khá chặt chẽ
- Nếu	0,7 ≤ R <0,9	- Nếu	0,5 ≤ R^2 <0,8	Tương quan chặt chẽ
- Nếu	0,9 ≤ R	- Nếu	0,8 ≤ R^2	Tương quan rất chặt chẽ

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	9.970	.491		20.304	.000
	Thu nhap nam (trieu)	1.162E-04	.000	.514	8.433	.000

a. Dependent Variable: Nam lam viec

Bảng coefficient cho phép chúng ta kiểm định các hệ số góc trong mô hình, ta có t_1 = 8,433 và p-value = 0,000<0,05 nên ta khẳng định tồn tại mối quan hệ giữa hai biến với hệ số góc b_1=0,00011 có nghĩa là khi tăng mỗi năm làm việc, thu nhập hàng năm tang 110 ngàn đồng. Ta có thể thành lập được phương trình hồi quy như sau:

$$y_i = 9.870 + 0,00011x_i + e$$

KIỂM ĐỊNH CHI BÌNH PHƯƠNG VỀ TÍNH ĐỘC LẬP HAY PHỤ THUỘC GIỮA HAI BIẾN (CROSSTABS)

Kiểm định phân phối (kiểm định sự phù hợp)

Tình huống: Trong một nghiên cứu ước tính của bộ Y tế, người ta mong muốn kiểm tra giả thuyết rằng tần suất sử dụng dịch vụ bệnh viện của các ngày trong tuần là như nhau và giảm 25% vào cuối tuần. Một mẫu gồm 52 000 bệnh nhân có phân phối sau:

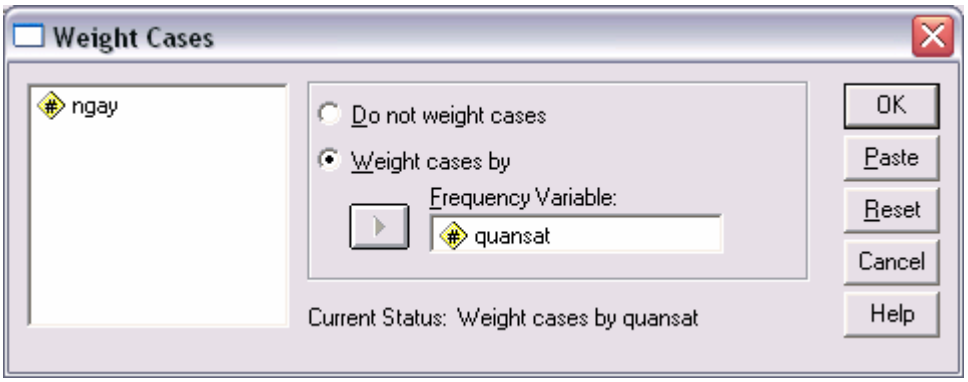
Ngày	Số bệnh nhân (quan sát)	Số bệnh nhân (lí thuyết)
Thứ hai	8623	8000
Thứ ba	8308	8000
Thứ tư	8420	8000
Thứ năm	9032	8000
Thứ sáu	8754	8000
Thứ bảy	4361	6000
Chủ nhật	4502	6000
	52000	**52000**

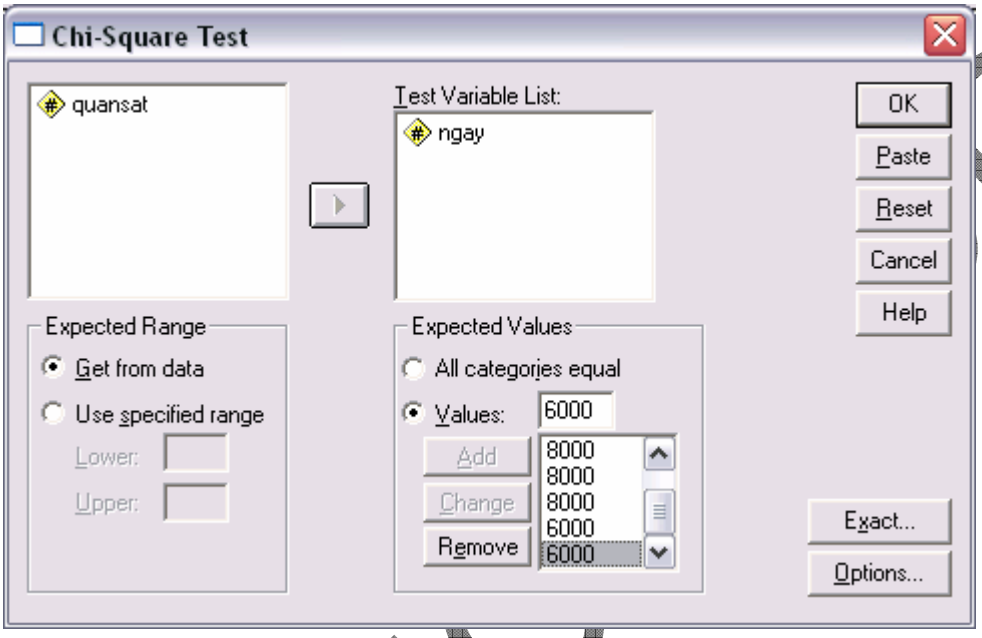
Khi đó, giả thiết và đối thiết:

H_0: Nhu cầu khám chữa bệnh là như nhau ở tất cả các ngày trong tuần và giảm 25% vào cuối tuần

H_1: Nhu cầu này có một dạng phân phối khác

Lê Văn Huy, Ph.D Candidate (2007) – Danang University of Economics, Email: *huy.lv@due.edu.vn*





Kiểm định chi bình phương về tính chất độc lập hay phụ thuộc (kiểm định hàng cột hay kiểm định mối quan hệ giữa hai biến biểu danh)

Người ta dùng kiểm định Chi bình phương để kiểm định sự kết hợp giữa bai biến (biểu danh hoặc thứ tự). Có một số chú ý như sau:

- χ^2 được thiết lập để xác định có hay không một mối liên hệ giữa hai biến, nhưng nó không chỉ ra được cường độ của mối liên hệ đó. Trong trường hợp này, cần sử dụng các đo lường kết hợp.
- χ^2 cho phép tìm ra những mối liên hệ phi tuyến tính giữa hai biến.
- Với kiểm định Chi bình phương, ta thành lập được các bảng chéo. Hệ số V Cramer được áp dụng cho tất cả các loại bảng chéo với k là chiều bé nhất của bảng chéo. Cường độ của nó biến động từ 0 đến 1.

$$V = \sqrt{\frac{\chi^2}{n(k-1)}}$$

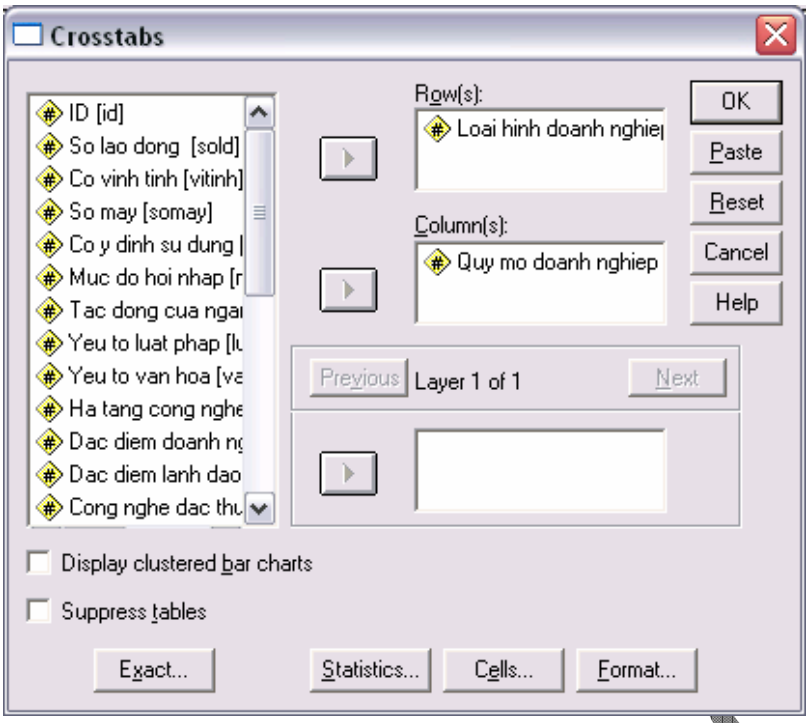
Giả sử ta chọn phân tích tính độc lập giữa hai biến định tính quy mô doanh nghiệp (quymo) và loại hình kinh doanh (loaihinh). Các bước tiến hành như sau:

H_0: Hai biến quy mô doanh nghiệp và loại hình doanh kinh độc lập với nhau trên tổng thể

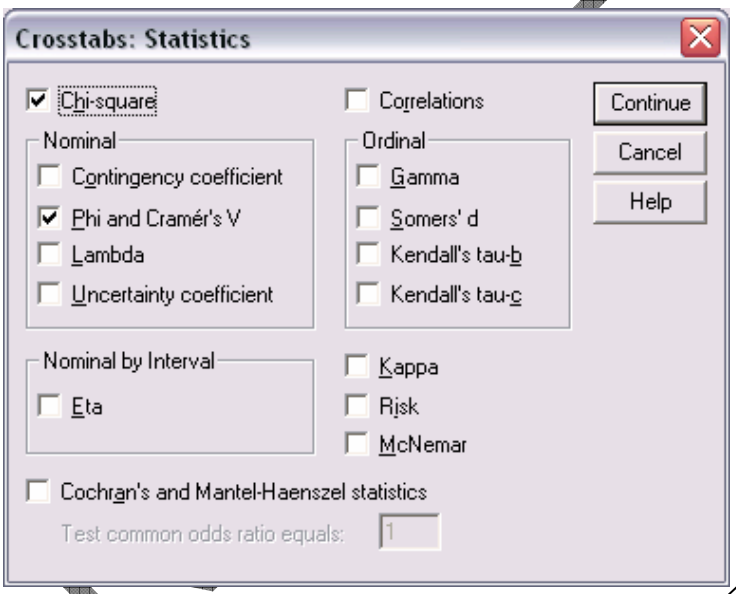
H_1: Hai biến quy mô doanh nghiệp và loại hình doanh kinh phụ thuộc với nhau trên tổng thể

Lê Văn Huy, Ph.D Candidate (2007) – Danang University of Economics, Email: *huy.lv@due.edu.vn*

☞ Vào **Descriptives statistics – Crosstab** chọn các biến vào các ô tương ứng



☞ Bấm Statistics để thiết lập các thống kê



☞ Bấm **Cells** để thiết lập các tỷ lệ phần trăm theo dòng, cột hay tổng cộng

Giá trị kiểm định Chi bình phương

Giá trị kiểm định p-value

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	38.665[a]	2	.000
Likelihood Ratio	50.910	2	.000
Linear-by-Linear Association	36.280	1	.000
N of Valid Cases	104		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 12.92.

Lê Văn Huy, Ph.D Candidate (2007) – Danang University of Economics, Email: *huy.lv@due.edu.vn*

Loai hinh doanh nghiep * Quy mo doanh nghiep Crosstabulation

			Quy mo doanh nghiep		
			vua va nho	lon	Total
Loai hinh doanh nghiep	Dich vu thuong mai	Count	11	26	37
		Expected Count	22.1	14.9	37.0
		% of Total	10.6%	25.0%	35.6%
	Xay dung	Count	16	16	32
		Expected Count	19.1	12.9	32.0
		% of Total	15.4%	15.4%	30.8%
	Cong nghiep	Count	35	0	35
		Expected Count	20.9	14.1	35.0
		% of Total	33.7%	.0%	33.7%
Total		Count	62	42	104
		Expected Count	62.0	42.0	104.0
		% of Total	59.6%	40.4%	100.0%

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.610	.000
	Cramer's V	.610	.000
N of Valid Cases		104	

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Trong kiểm này, ta thấy giá trị Chi bình phương = 38,665 và p-value=0,000<0,05 nên ta bác bỏ H_0 và chấp nhận H_1 tức hai biến phụ thuộc lẫn nhau trên tổng thể.

Hệ số Phi = 0,61 khẳng định mối quan hệ giữa hai biến này khá chặt chẽ.

KIỂM ĐỊNH PHI THAM SỐ

Kiểm định hai mẫu phụ thuộc (Wilcoxon, kiểm định dấu, kiểm định Nemar)

Với ví dụ về đánh giá hai loại kem ở trên, ta cógiả thiết:

Với giả thiết và đối thiết là:

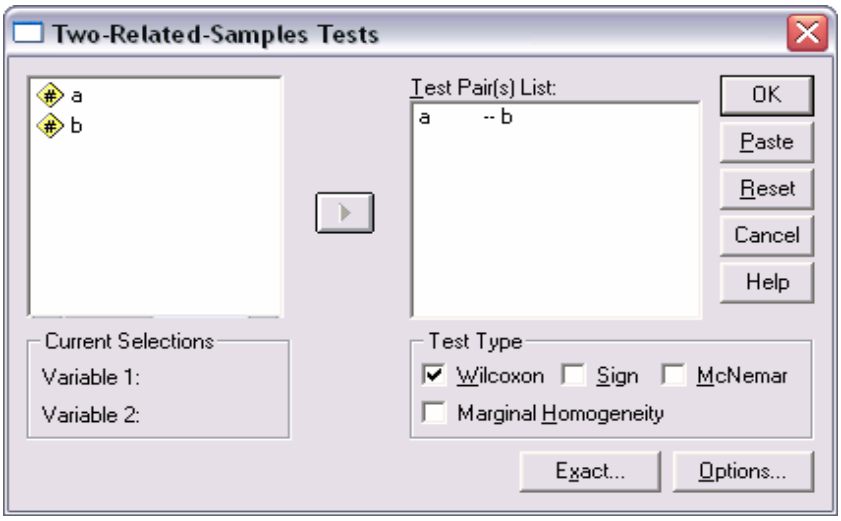
H_0: Không có sự khác biệt trong mức độ ưa chuộng giữa A, B trong tổng thể

H_1: Có sự khác biệt trong mức độ ưa chuộng giữa A, B trong tổng thể

Các bước thực hiện như sau:

☞ Vào **Analyze** – **Nonparametric Tests - 2 Related Samples**

Lê Văn Huy, Ph.D Candidate (2007) – Danang University of Economics, Email: *huy.lv@due.edu.vn*



☞ Kết quả thu được:

Ranks

		N	Mean Rank	Sum of Ranks
B - A	Negative Ranks	2[a]	1.50	3.00
	Positive Ranks	5[b]	5.00	25.00
	Ties	2[c]		
	Total	9		

a. B < A

b. B > A

c. A = B

Test Statistics [b]

	B - A
Z	-1.876[a]
Asymp. Sig. (2-tailed)	.061

a. Based on negative ranks.

b. Wilcoxon Signed Ranks Test

☞ Nhìn vào bảng trên ta có thể dễ dàng diễn giải dữ liệu, với Z = -1,876 và p-value=0,61>0,05 nên ta chưa có cơ sở để bác bỏ H_0 tức chưa có cơ sở để chấp nhận H_1 hay chưa có cơ sở để khẳng định có sự khác biệt trong mức độ ưa chuộng giữa A, B trong tổng thể.

Chú ý: Kiểm định dấu và Nemar có thể thực hiện tương tự

Kiểm định nhiều hơn hai mẫu phụ thuộc (Friedman, Kendall's W, Cochran's Q)

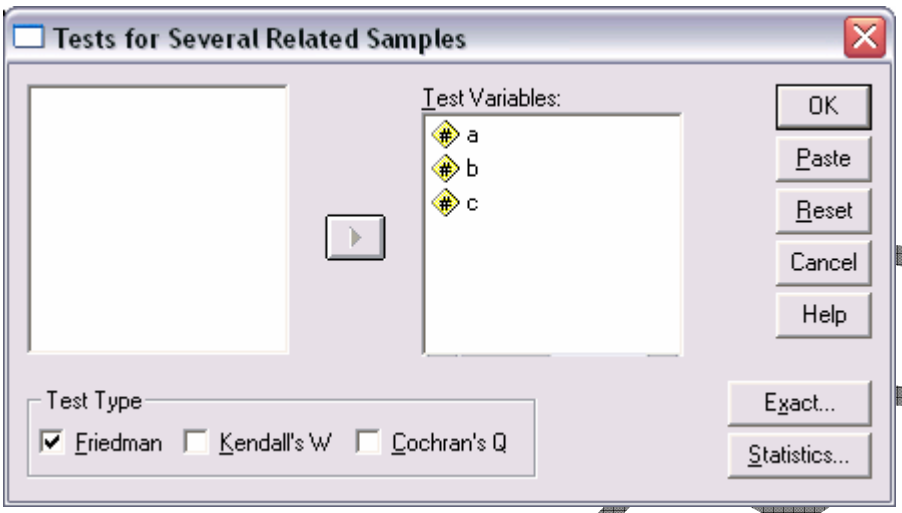
Trong trường hợp giống như ví dụ ở trường hợp kiểm định wilcoxon, nhưng bây giờ ta có 3 sản phẩm A, B, C, khi đó

KH	*Kem A*	*Kem B*	*Kem C*
1	*4*	*3*	*5*
2	*5*	*5*	*5*
3	*2*	*5*	*5*

Lê Văn Huy, Ph.D Candidate (2007) – Danang University of Economics, Email: *huy.lv@due.edu.vn*

4	3	2	5
5	3	5	5
6	1	5	5
7	3	3	5
8	2	5	5
9	2	5	5

☞ Vào **Analyze – Nonparametric Test – K Related Samples** chọn các biến vào phân tích



☞ Kết quả:

Ranks

	Mean Rank
A	1.39
B	2.00
C	2.61

Test Statistics [a]

N	9
Chi-Square	9.308
df	2
Asymp. Sig.	.010

a. Friedman Test

☞ Với Chi bình phương = 9,308 và p-value=0,01<0,05 nên ta bác bỏ H_0 tức chấp nhận H_1 hay đã có sự khác biệt trong mức độ ưa chuộng giữa A, B, C trong tổng thể.

Kiểm định cho hai mẫu độc lập (Mann-Whitney U)

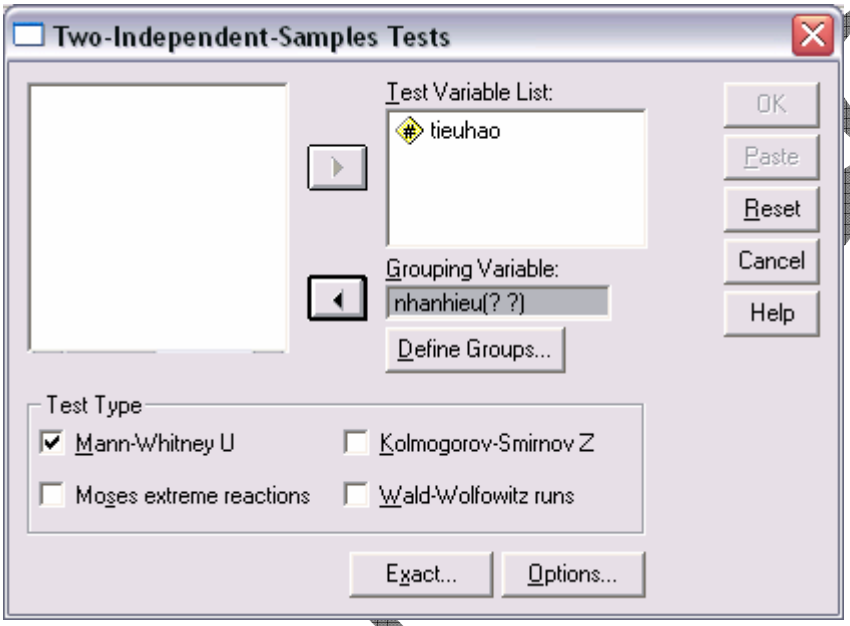
Tình huống: Có hai loại máy nổ Toshiba và Yamaha đang tiêu thụ tại Việt Nam, một nhà phân phối muốn kiểm tra mức độ tiêu hao nguyên vật liệu của hai loại sản phẩm này.

Nhà phân phối gặp các khách hàng sử dụng hai loại sản phẩm, tiến hành điều tra mức tiêu hao nguyên vật liệu, tổng số khách hàng điều tra là 18 người trong đó 10 người sử dụng sản phẩm Toshiba và 10 người sử dụng sản phẩm Yamaha, kết quả thu được như sau:

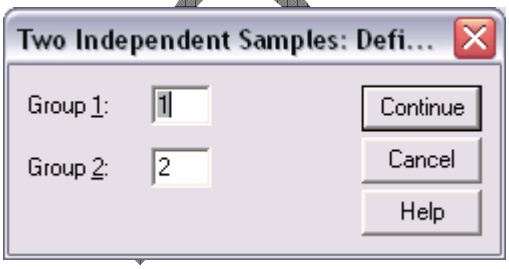
Lê Văn Huy, Ph.D Candidate (2007) – Danang University of Economics, Email: huy.lv@due.edu.vn

KH	*Toshiba*	*Yamaha*
1	4000	4200
2	3800	4300
3	4600	3400
4	4300	3500
5	5000	3800
6	5300	4200
7	4900	4300
8	4700	3400
9	4000	
10	5200	

☞ Vào **Analyze – Nonparametric Test – 2 Independent Samples** chọn các biến vào phân tích



☞ Nhấn **Grouping Define** để định nghĩa các biến



☞ Kết quả như sau

Ranks

	NHANHIEU	N	Mean Rank	Sum of Ranks
TIEUHAO	Toshiba	10	12.15	121.50
	Yamaha	8	6.19	49.50
	Total	18		

Lê Văn Huy, Ph.D Candidate (2007) – Danang University of Economics, Email: *huy.lv@due.edu.vn*

Test Statistics [b]

	TIEUHAO
Mann-Whitney U	13.500
Wilcoxon W	49.500
Z	-2.364
Asymp. Sig. (2-tailed)	.018
Exact Sig. [2*(1-tailed Sig.)]	.016[a]

a. Not corrected for ties.

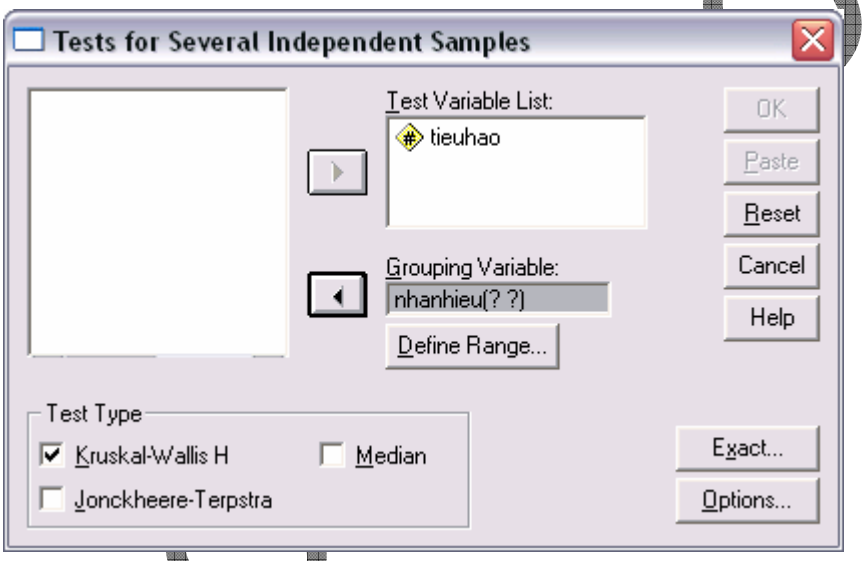
b. Grouping Variable: NHANHIEU

Ta thấy giá trị Mann-Whitney U = 13,5, giá trị Z = -2,363 và p-value=0,18 nên ta bác bỏ H_0, chấp nhận H_1 tức là có sự khác nhau về mức tiêu hao nhiên liệu trung bình của hai loại sản phẩm là khác nhau.

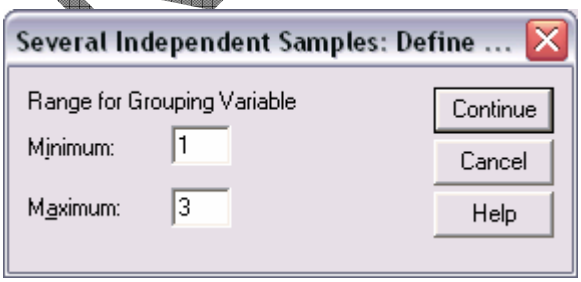
Kiểm định nhiều hơn hai mẫu độc lập (Kruskal-Wallis H)

Giả sử như chúng ta có 3 nhóm sản phẩm (thêm một sản phẩm của hãng Sonix), cách thức thực hiện như sau:

☞ Vào **Analyze – Nonparametric Test – K Independent Samples** chọn các biến vào phân tích



☞ Vào **Grouping Variable** để định nghĩa biến, hiện tại chúng ta có 3 biến, chúng ta đặt giá trị ở maximum là 3, nếu chúng ta so sánh 2 biến thì chúng ta định số 2 (giá trị 1, 2, hay 3 phụ thuộc vào các định nghĩa value label của biến nhanhieu.



Lê Văn Huy, Ph.D Candidate (2007) – Danang University of Economics, Email: *huy.lv@due.edu.vn*

Ranks

	NHANHIEU	N	Mean Rank
TIEUHAO	Toshiba	10	18.55
	Yamaha	8	9.63
	3	8	11.06
	Total	26	

Test Statistics [a,b]

	TIEUHAO
Chi-Square	7.318
df	2
Asymp. Sig.	.026

a. Kruskal Wallis Test

b. Grouping Variable: NHANHIEU

Ta thấy giá trị Chi bình phương = 7,318 và p-value=0,026 nên ta bác bỏ H_0, chấp nhận H_1 tức là có sự khác nhau về mức tiêu hao nhiên liệu trung bình của ba loại sản phẩm là khác nhau.

Lê Văn Huy, Ph.D Candidate (2007) – Danang University of Economics, Email: *huy.lv@due.edu.vn*