PHƯƠNG PHÁP XỬ LÝ VÀ PHÂN TÍCH SỐ LIỆU NGHIÊN CỨU

TS.BS Võ Bảo Dũng

Xử lý và phân tích số liệu hay dữ liệu nghiên cứu là một trong các bước cơ bản của một nghiên cứu, bao gồm xác định vấn đề nghiên cứu; thu thập số liệu; xử lý số liệu; phân tích số liệu và báo cáo kết quả. Xác định rõ vấn đề nghiên cứu giúp việc thu thập số liệu được nhanh chóng và chính xác hơn. Để có cơ sở phân tích số liệu tốt thì trong quá trình thu thập số liệu phải xác định trước các yêu cầu của phân tích để có thể thu thập đủ và đúng số liệu như mong muốn.

Điều cốt lõi của phân tích số liệu là suy diễn thống kê, nghĩa là mở rộng những hiểu biết từ một mẫu ngẫu nhiên thành hiểu biết về tổng thể, hay còn gọi là suy diễn quy nạp. Muốn có được các suy diễn này phải phân tích số liệu dựa vào các test thống kê để đảm bảo độ tin cậy của các suy diễn. Bản thân số liệu chỉ là các số liệu thô, qua xử lý phân tích trở thành thông tin và sau đó trở thành tri thức. Đây chính là điều mà tất cả các nghiên cứu đều mong muốn.

Trong khuôn khổ của bài viết này, tác giả muốn trình bày với người đọc một số khái niệm và kỹ thuật cơ bản khi xử lý và phân tích số liệu. Bên cạnh đó, tác giả giới thiệu một số thao tác cơ bản khi xử lý số liệu trên phần mềm SPSS, một phần mềm xử lý thông kê được sử dụng nhiều nhất hiện nay.

1. Xử lý số liệu nghiên cứu

Ngày nay, hầu hết các nghiên cứu đều xử lý số liệu trên các phần mềm máy tính. Do vậy, việc xử lý số liệu phải qua các bước sau:

- Mã hóa số liệu: Các số liệu định tính (biến định tính) cần được chuyển đổi (mã hóa) thành các con số. Các số liệu định lượng thì không cần mã hóa.
- Nhập liệu: Số liệu được nhập và lưu trữ vào file dữ liệu. Cần phải thiết kế khung file số liệu thuận tiện cho việc nhập liệu.
- Hiệu chỉnh: Là kiểm tra và phát hiện những sai sót trong quá trình nhập số liệu từ bảng số liệu ghi tay vào file số liệu trên máy tính.

2. Phân loại các số liệu (biến số) trong nghiên cứu

Có 2 loại biến số chính trong hầu hết các nghiên cứu đố là biến số định tính và biến số định lượng.

- Biến định tính: là loại biến số phản ảnh tính chất, sự hơn kém. Có thể biểu diễn dưới dạng định danh (ví dụ: nam/nữ) hay thứ bậc (tốt/khá/trung bình/yếu)…Đối với loại biến số này ta không tính được giá trị trung bình của số liệu.
- Biến định lượng: Thường được biểu diễn bằng các con số. Các con số này có thể ở dưới dạng biến thiên liên tục (ví dụ: huyết áp của bệnh nhân theo thời gian) hoặc rời rạc (ví dụ: chiều cao, cân nặng của người bệnh lúc vào viện). Dạng

biến này cho phép chúng ta tính được giá trị trung bình của biến. Cần lưu ý là tất cả các biến định lượng đều phải có đơn vị tính (mmHg, mmol/L, mg%, Kg…)

3. Một số nguyên tắc chọn test thống kê trong phân tích số liệu nghiên cứu

Khi chọn một test thống kê cần cân nhắc các yếu tố sau:

- Mục tiêu nghiên cứu: Đo lường sự khác nhau hay mối tương quan giữa các biến số.
- Số nhóm nghiên cứu: một nhóm, hai nhóm hay trên 2 nhóm.
- Cỡ mẫu nghiên cứu: nhỏ hay lớn.
- Bản chất của số liệu, loại biến số: định tính hay định lượng.
- Phân bố mẫu: phân bố chuẩn hay không chuẩn.
- Loại quan sát: mẫu độc lập hay ghép cặp.

Có 2 loại test thống kê căn bản đó là test tham số (parametric test) và test phi tham số (non-parametric test).

Test tham số chỉ dùng cho các nghiên cứu có cỡ mẫu lớn (> 30), phân bố của quần thể là phân bố chuẩn, thường dùng với các biến định lượng, các test thường dùng là t test, ANOVA, tương quan Pearson, hồi quy tuyến tính…

Test phi tham số có thể áp dụng cho các mẫu nghiên cứu lớn và nhỏ, phân bố của quần thể không biết được hoặc có phân bố không chuẩn, có thể dùng cho cả biến danh mục hoặc thứ hạng (định tính), các test thường dùng là Fisher test, Median test, χ^2 test, Kruskal-Wallis one way ANOVA test, tương quan thứ hạng Spearman…

4. Phân tích số liệu nghiên cứu

4.1. Mô tả các biến số

- Tỷ lệ: đối với các biến định tính.

Từ tỷ lệ có thể ước lượng từ mẫu ra quần thể nghiên cứu với các phép ước lượng điểm, ước lượng khoảng hoặc kiểm định giả thuyết cho tỷ lệ cuả một nhóm hay nhiều nhóm.

- Giá trị trung bình (X), trung vị , độ lệch chuẩn (SD), giá trị lớn nhất, giá trị nhỏ nhất: đối với các biến định lượng. Tương tự như với biến định tính, từ các giá trị trung bình của mẫu có thể ước lượng điểm, ước lượng khoảng ra các giá trị của quần thể.

4.2. Phân tích sự khác biệt

- So sánh 2 tỷ lệ.
- So sánh 3 tỷ lệ hoặc hơn.
- So sánh 2 giá trị trung bình.
- So sánh 3 giá trị trung bình hoặc hơn.

4.3. Phân tích mối liên quan giữa các biến số

- Tương quan giữa 2 biến định tính:

+ Tỷ suất chênh hiện mắc (POR: Prevalence Odd Ratio): trong nghiên cứu ngang.

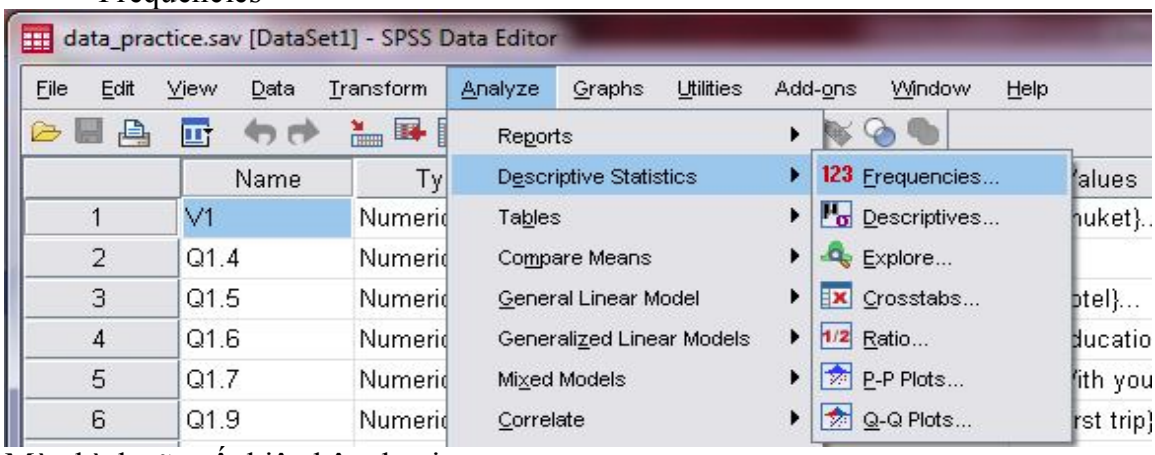
+ Tỷ suất chênh OR: trong nghiên cứu bệnh chứng không ghép cặp.
+ Nguy cơ tương đối RR (Relative Risk): trong nghiên cứu thuần tập.

- Tương quan giữa 2 biến định lượng:
  + Hệ số tương quan r.
  + Phương trình hồi quy tuyến tính: Y = a + bX
- Tương quan giữa 3 biến định tính trở lên: phân tích tầng.
- Tương quan hồi quy tuyến tính bội

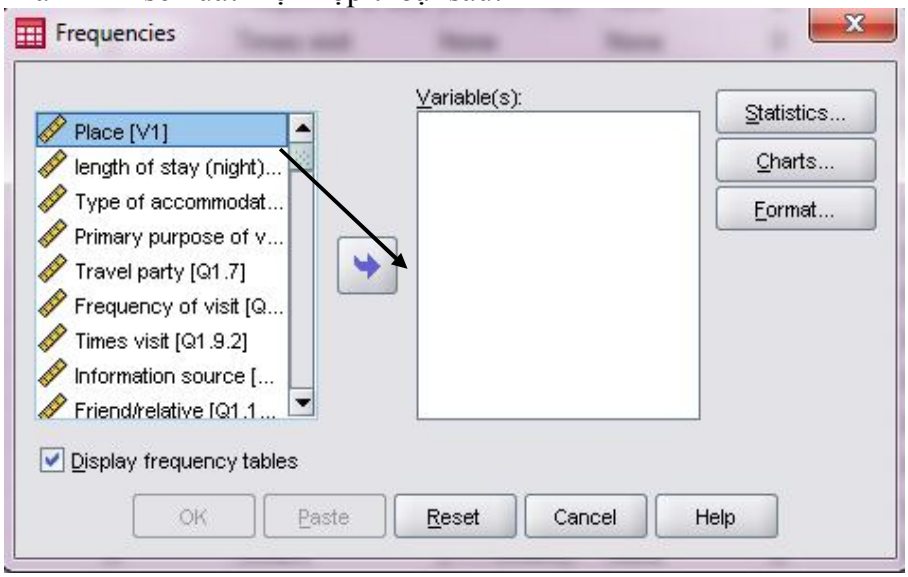
5. Phân tích số liệu với phần mềm SPSS: Một số thao tác cơ bản

CÁCH THỨC TIẾN HÀNH LỆNH FREQUENCIES
(Tính tần số)

1. Sau khi mở file dữ liệu, vào menu Analyze → Descriptive Statistics → Frequencies



Màn hình sẽ xuất hiện hộp thoại sau:



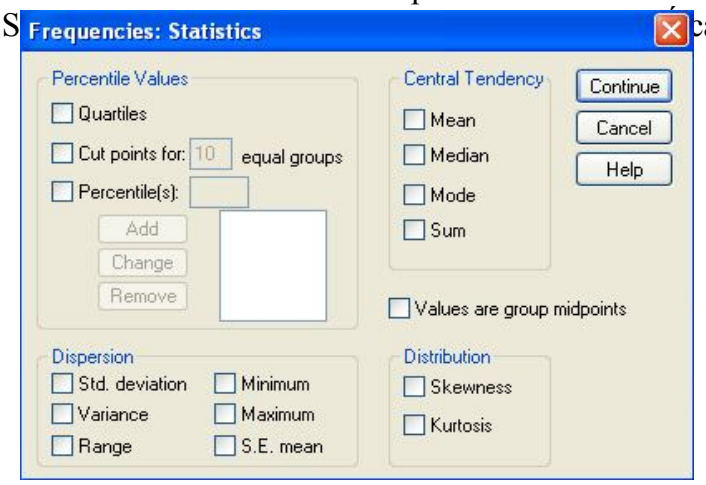
← Vẽ biểu đồ

2. Chọn biến muốn tính tần số (biến Place V1) bằng cách click chuột vào tên biến rồi đưa sang khung Variable(s).
3. Click Ok. Trường hợp muốn vẽ biểu đồ thực hiện thêm bước 4 trước khi click Ok.
4. Để vẽ biểu đồ click chuột vào ô Charts…. Chọn dạng biểu đồ ở Chart type, chọn giá trị thể hiện trên biểu đồ là số đếm (frequencies) hay phần trăm (percentages). Click Continue để trở lại hộp thoại Frequencies → Ok để thực hiện lệnh.

TÍNH TRỊ TRUNG BÌNH (2 CÁCH)

CÁCH 1. DÙNG LỆNH FREQUENCY

1. Vào menu Analyze → Descriptive Statistics → Frequencies. Đưa biến cần tính trị trung bình vào ô variables như bước 1 và 2 ở phần trước.
2. Click chọn thẻ S… cần thiết:

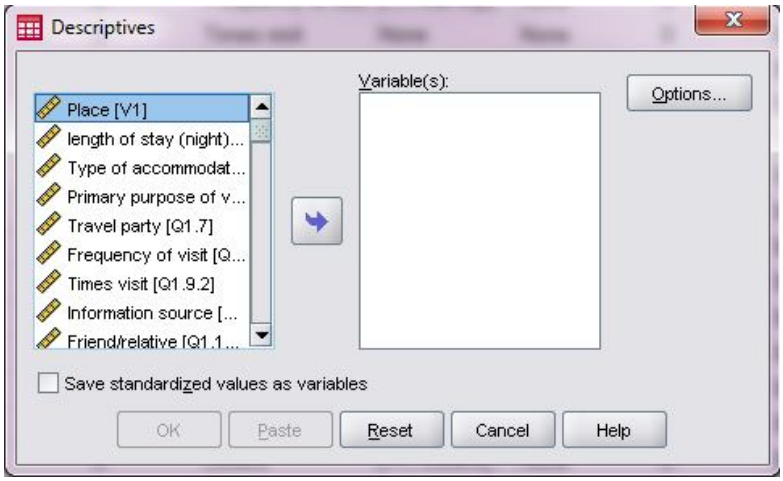


3. Ý nghĩa một số thông số thông dụng:

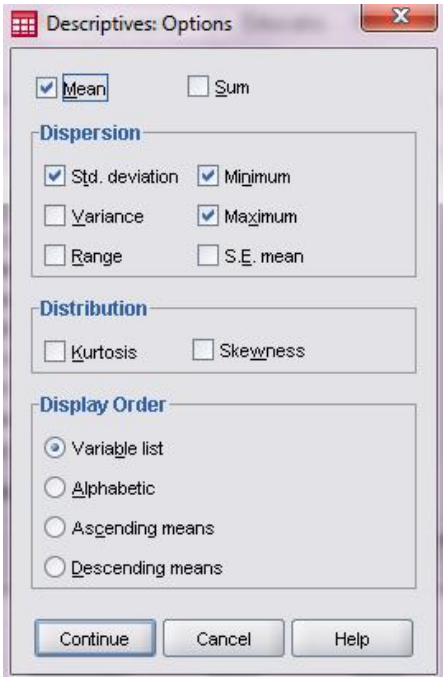
Mean:	trung bình cộng
Sum:	tổng cộng (cộng tất cả các giá tị trong tập dữ liệu quan sát)
Std. Deviation:	độ lệch chuẩn
Minimum:	giá trị nhỏ nhất
Maximum:	giá trị lớn nhất
S.E. mean:	sai số chuẩn khi ước lượng trị trung bình

CÁCH 2. DÙNG LỆNH DESCRIPTIVES
(Tính điểm trung bình)

1. Vào menu Analyze → Descriptive Statistics → Descriptives…, xuất hiện hộp thoại



2. Chọn 1 hay nhiều biến (định lượng) muốn tính điểm trung bình đưa vào khung Variable(s).
3. Click vào ô Options… để xuất hiện hộp thoại Descriptive Options. Chọn các đại lượng thống kê muốn tính toán bằng cách click vào ô vuông cần thiết.
4. Chọn cách sắp xếp kết quả tính toán theo thứ tự danh sách biến (Variable list), thứ tự Alphabetic của nhãn biến, thứ tự tăng dần (Ascending list), và thứ tự giảm dần (Descending list).
5. Click Continue để trở về hộp thoại Descriptive → Ok để thực hiện lệnh.

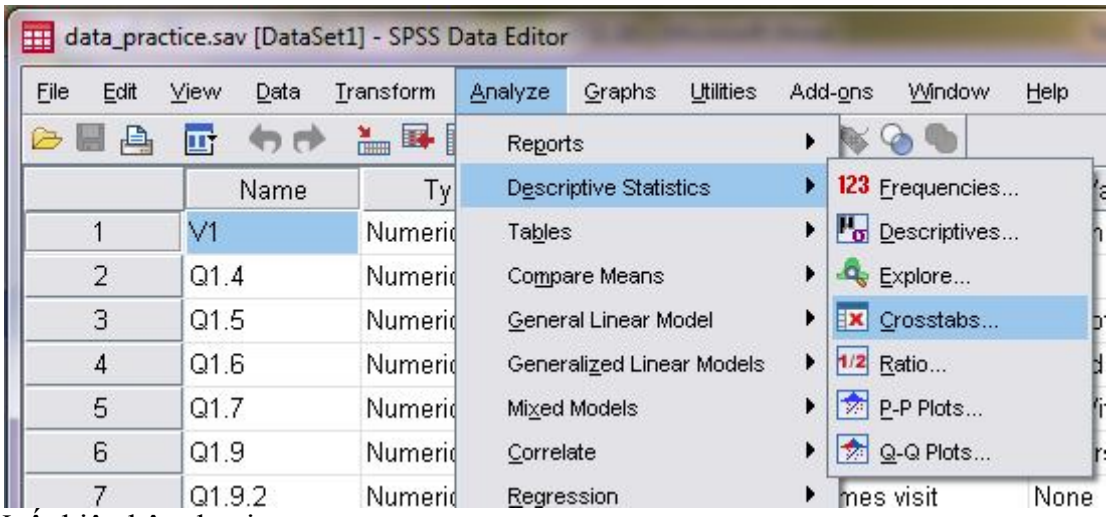


CÁCH THỨC TIẾN HÀNH PHÂN TÍCH BẢNG CHÉO

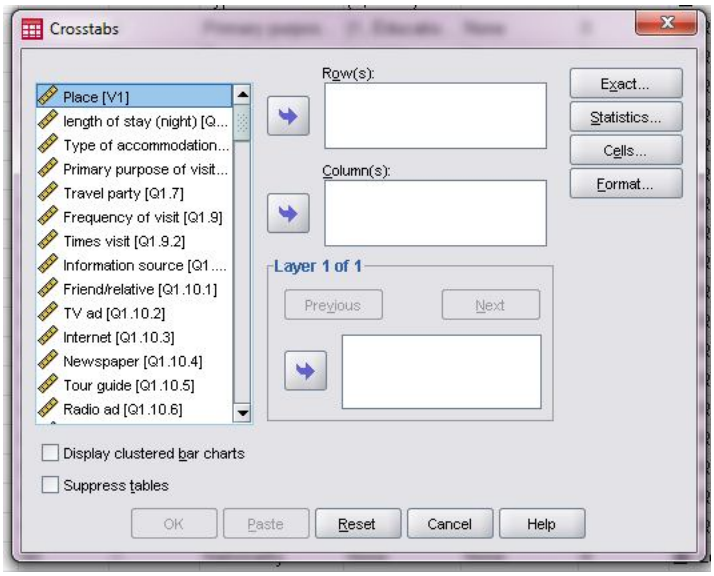
(CROSSTABULATION)

Phân tích bảng chéo dùng để kiểm định mối quan hệ giữa các biến định tính với nhau bằng cách dùng kiểm định Chi – bình phương (Chi-square). Cách thức tiến hành với SPSS như sau:

1. Vào menu Analyze → Descriptive Statistics → Crosstabs…,

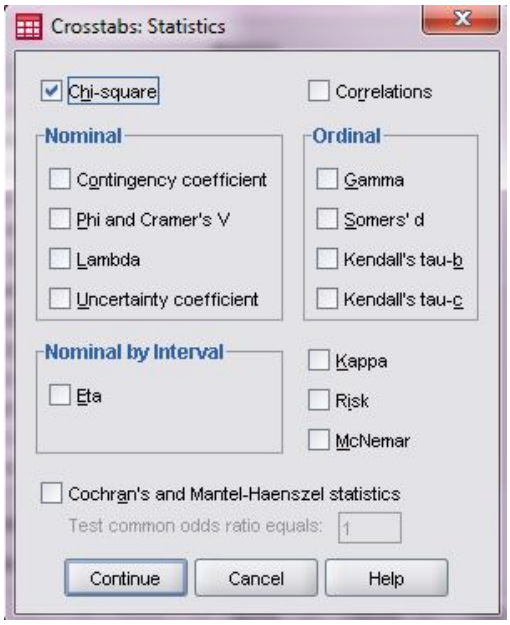


2. Xuất hiện hộp thoại sau:



3. Chọn và đưa các biến vào khung Row(s) (dòng) và Column(s) (cột) và Layer 1 of 1 (đối với trường hợp trên 2 biến).

4. Click vào ô Statistics, xuất hiện hộp thoại sau:



5. Chọn các kiểm định cần thiết. Trong trường hợp này ta dùng kiểm định Chi – bình phương (Chi-square).
   - Các kiểm định ở ô Norminal dùng để kiểm định mối liên hệ giữa các biến biểu danh.
   - Các kiểm định ở ô Ordinal dùng để kiểm định mối liên hệ giữa các biến thứ tự.
6. Click vào continue để trở lại hộp thoại Crosstabs → Click vào ô Cells, hộp thoại sau xuất hiện:

7. Ở ô Counts chọn Observed (thể hiện tần số quan sát). Trong

trường hợp muốn thể hiện tần số mong đợi chọn Expected.

8. Chọn cách thể hiện phần trăm theo dòng hay theo cột ở ô Percentages.
9. Click Continue để trở lại hộp thoại Crosstabs → Ok để thực hiện lệnh.

CÁCH ĐỌC KẾT QUẢ KIỂM ĐỊNH

Khi thực hiện kiểm định, ta có 2 giả thuyết.

H_0: không có mối quan hệ giữa các biến.

H_1: có mối quan hệ giữa các biến.

Để kết luận là chấp nhận hay bác bỏ giả thuyết H_0, ta sẽ dùng các kiểm định phù hợp.

Dựa vào giá trị P (p-value) (SPSS viết tắt p-value là sig.) để kết luận là chấp nhận hay bác bỏ giả thuyết H_0

p-value (sig.) ≤ α (mức ý nghĩa) → bác bỏ giả thuyết H_0. Có nghĩa là có mối quan hệ có ý nghĩa giữa các biến cần kiểm định.

p-value (sig.) > α (mức ý nghĩa) → chấp nhận H_0. Không có mối quan hệ giữa các biến cần kiểm định.

ĐỐI VỚI KIỂM ĐỊNH CHI – BÌNH PHƯƠNG

Hàng đầu tiên của bảng Chi-square tests thể hiện giá trị P

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	16.217[a]	8	.039
Likelihood Ratio	18.708	8	.017
Linear-by-Linear Association	.202	1	.653
N of Valid Cases	511		

p-value

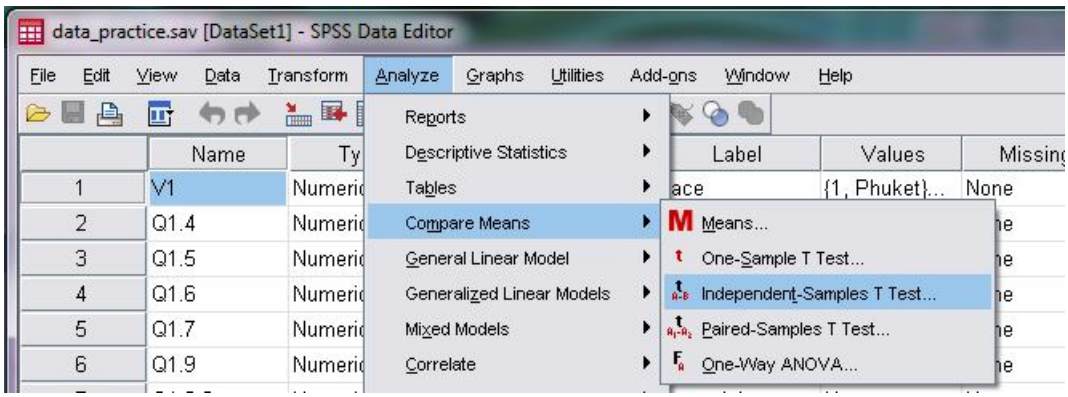
a. 8 cells (44.4%) have expected count less than 5. The minimum expected count is 1.69.

Cuối bảng Chi-Square tests SPSS sẽ đưa ra dòng thông báo cho biết % số ô có tần suất mong đợi dưới 5. Kiểm định Chi-bình phương chỉ có ý nghĩa khi số quan sát đủ lớn, nếu có quá 20% số ô trong bảng chéo có tần số lý thuyết nhỏ hơn 5 thì giá trị chi-bình phương không còn đáng tin cậy.

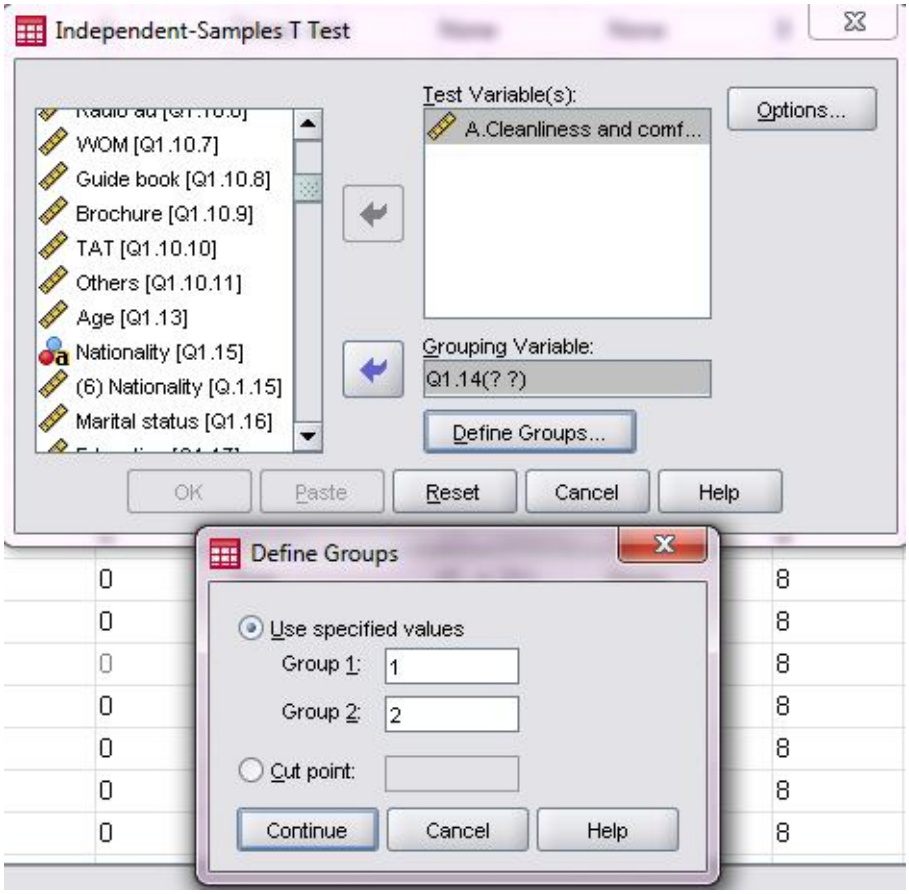
Trong ví dụ trên có đến 44.4% số ô có tần số mong đợi dưới 5, biện pháp cho trường hợp này là ta sẽ gom các biểu hiện trên các biến lại để tăng số quan sát trong mỗi nhóm.

CÁCH THỨC TIẾN HÀNH KIỂM ĐỊNH GIẢ THUYẾT VỀ TRỊ TRUNG BÌNH CỦA 2 TỔNG THỂ ĐỘC LẬP (Independent Samples T-test)

1. Vào menu Analyze → Compare Means → Independent-samples T-test



2. Chọn biến định lượng cần kiểm định trị trung bình đưa vào khung Test Variable(s). Chọn biến định tính chia số quan sát thành 2 nhóm mẫu để so sánh giữa 2 nhóm này với nhau đưa vào khung Grouping Variable.
3. Chọn Define Groups… để nhập mã số của 2 nhóm. Click Continue để trở lại hộp thoại chính → Click Ok để thực hiện lệnh



Trong kiểm định Independent-samples T-test, ta cần dựa vào kết quả kiểm định sự bằng nhau của 2 phương sai tổng thể (kiểm định Levene). Phương sai diễn tả mức độ đồng đều hoặc không đồng đều (độ phân tán) của dữ liệu quan sát.

Independent Samples Test

			A.Cleanliness and comfort of room	
			Equal variances assumed	Equal variances not assumed
Levene's Test for Equality of Variances	F		.138	
	Sig.		.710	
t-test for Equality of Means	t		-3.066	-3.040
	df		509	448.100
	Sig. (2-tailed)		.002	.003
	Mean Difference		-.231	-.231
	Std. Error Difference		.075	.076
	95% Confidence Interval of the Difference	Lower	-.379	-.380
		Upper	-.083	-.082

Nếu giá trị Sig. trong kiểm định Levene (kiểm định F) < 0.05 thì phương sai của 2 tổng thể khác nhau, ta sử dụng kết quả kiểm định t ở dòng Equal variances not assumed.

Nếu Sig. ≥ 0.05 thì phương sai của 2 tổng thể không khác nhau, ta sử dụng kết quả kiểm định t ở dòng Equal variances assumed.

Trong VD trên Sig. của kiểm định F = 0.71 > 0.05 ➔ chấp nhận giả thuyết H_0 không có sự khác nhau về phương sai của 2 tổng thể ➔ sử dụng kết quả ở dòng Equal variances assumed.

Nếu Sig. của kiểm định t ≤ α (mức ý nghĩa) → có sự phác biệt có ý nghĩa về trung bình của 2 tổng thể.

Nếu Sig. > α (mức ý nghĩa) → không có sự khác biệt có ý nghĩa về trung bình của 2 tổng thể.

Trong VD trên sig. = 0.002 < 0.05 ➔ có sự khác biệt có ý nghĩa về trung bình của 2 tổng thể.

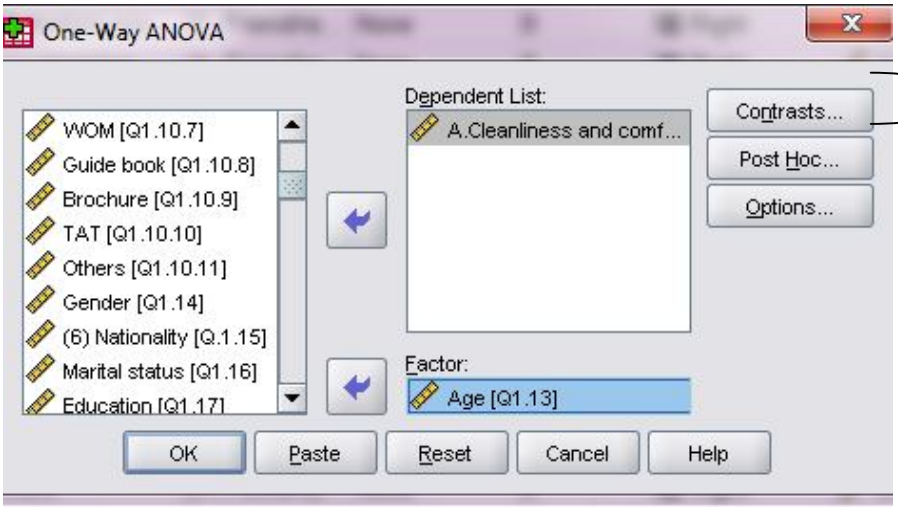
CÁCH THỨC TIẾN HÀNH PHÂN TÍCH PHƯƠNG SAI
(ANOVA – Analysis of Variance)

Phân tích phương sai ANOVA là phương pháp so sánh trị trung bình của 3 nhóm trở lên. Có 2 kỹ thuật phân tích phương sai: ANOVA 1 yếu tố (một biến yếu tố để phân loại các quan sát thành các nhóm khác nhau) và ANOVA nhiều yếu tố(2 hay nhiều biến để phân loại). Ở phần thực hành cơ bản chỉ đề cập đến phân tích phương sai 1 yếu tố (One-way ANOVA).

Một số giả định đối với phân tích phương sai một yếu tố:

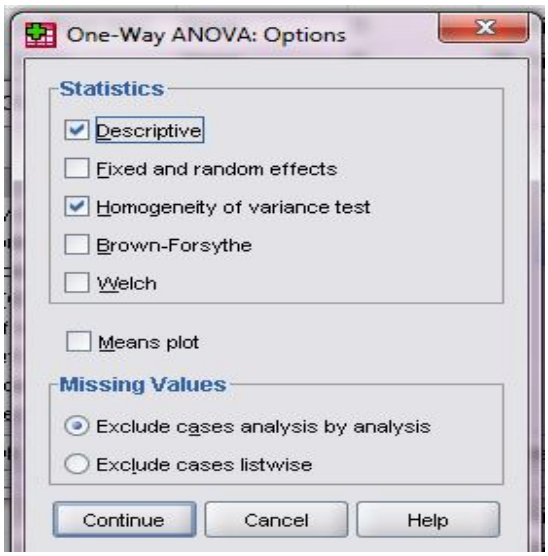
- Các nhóm so sánh phải độc lập và được chọn một cách ngẫu nhiên.
- Các nhóm so sánh phải có phân phối chuẩn hoặc cỡ mẫu phải đủ lớn để được xem như tiệm cận phân phối chuẩn.
- Phương sai của các nhóm so sánh phải đồng nhất.

1. Từ menu Analyze → Compare Means → One-Way ANOVA, xuất hiện hộp thoại sau:



2 kỹ thuật dùng để kiểm định sâu ANOVA

2. Đưa biến định lượng (trị trung bình) vào khung Dependent list.
   Đưa biến phân loại xác định các nhóm cần so sánh với nhau vào khung Factor.
3. Click vào nút Option để mở hộp thoại One-Way ANOVA Options.
   Trong hộp thoại One-way ANOVA Options:
   - Click chọn ô Descriptive để tính đại lượng thống kê mô tả (tính trị trung bình) theo từng nhóm so sánh.
   - Click chọn ô Homogeneity of variance test để kiểm định sự bằng nhau của các phương sai nhóm (thực hiện kiểm định Levene).



4. Click chọn Continue để trở lại hộp thoại ban đầu → click Ok để thực hiện lệnh.

5. Dựa vào kết quả kiểm định ANOVA, nếu H_0 được chấp nhận thì kết luận không có sự khác biệt có ý nghĩa giữa các nhóm với nhau. Nếu H_0 bị bác bỏ → có sự khác biệt có ý nghĩa giữa các nhóm → trở lại hộp thoại One – way ANOVA để thực hiện kiểm định sâu ANOVA nhằm xác định cụ thể trung bình của nhóm nào khác với nhóm nào, nghĩa là tìm xem sự khác biệt của các nhóm xảy ra ở đâu.

6. Tuy nhiên có thể thực hiện kiểm định ANOVA và sâu ANOVA cùng lúc với nhau. Dựa vào sự chấp nhận hay bác bỏ giả thuyết H_0 để quan tâm hay không quan tâm đến kết quả kiểm định sâu ANOVA.

Tài liệu tham khảo:

1. Đỗ Hàm (2009), Phương pháp luận trong nghiên cứu khoa học y học, Trường Đại học Y khoa Thái Nguyên.
2. Phạm Văn Hiền (2009), Phương pháp tiếp cận nghiên cứu khoa học, http://pgo.hcmuaf.edu.vn/pvhien.
3. Nguyễn Văn Hộ, Nguyễn Đăng Bình (2004), Phương pháp luận nghiên cứu khoa học, Đại học Thái Nguyên.
4. Đinh Thanh Huề (2004), Phương pháp nghiên cứu khoa học, Trường Đại học Y Huế.
5. Hoàng Trọng, Chu Nguyễn Mộng Ngọc (2008), Phân tích dữ liệu nghiên cứu với SPSS, Trường Đại học Kinh tế TP Hồ Chí Minh.
6. Nguyễn Văn Tuấn (2009), Phương pháp nghiên cứu khoa học giáo dục, Trường Đại học SPKT TPHCM.