

XÂY DỰNG BIỂU ĐỒ HỘP (BOXPLOT) BẰNG PHẦN MỀM SPSS VÀ SỬ DỤNG TRONG NGHIÊN CỨU, GIẢNG DẠY ĐỊA LÝ KINH TẾ - XÃ HỘI

USING SPSS SOFTWARE TO BUILD BOX PLOT IN RESEARCHING AND TEACHING SOCIAL – ECONOMIC GEOGRAPHY

Trương Văn Cảnh

Trường Đại học Sư phạm – Đại học Đà Nẵng

Email: trvcanh1712@gmail.com

TÓM TẮT

Biểu đồ là công cụ rất quan trọng và phổ biến trong dạy học và nghiên cứu địa lý. Sử dụng và phổ biến các tính năng của SPSS trong dạy học và nghiên cứu địa lý là rất quan trọng, trong đó có việc sử dụng các dạng biểu đồ. Chúng ta có thể xây dựng dạng biểu đồ một cách nhanh chóng và tiện lợi bằng Excel. Tuy nhiên, đối với dạng biểu đồ hộp, bên cạnh việc xây dựng biểu đồ thì vấn đề xử lý, phân tích số liệu cũng rất quan trọng. Vì vậy, phần mềm SPSS được tác giả lựa chọn. Bài viết đã tập trung trình bày khá chi tiết về cách thức xây dựng và ý nghĩa biểu hiện của dạng biểu đồ hộp trong SPSS, đồng thời đưa ra ví dụ minh họa về việc vận dụng cụ thể trong phân tích một vấn đề của Địa lý kinh tế – xã hội.

Từ khóa: SPSS; biểu đồ; biểu đồ hộp; địa lý kinh tế – xã hội.

ABSTRACT

Charts are a very important and popular tool for teaching and researching geography. Using and applying features of SPSS to teaching and researching geography are essential, including the use of the charts. A chart can be built quickly and conveniently by Excel. Especially for the box charts, besides building a chart, processing and analyzing data are also important. Therefore, SPSS software is selected. The article focuses on how to build the box plot and the expression meaning of the box chart in SPSS, and gives illustrative examples of the application of the software to analyzing the economic – social geography problem.

Key words: SPSS; chart; box plot; economic – social geography.

1. Đặt vấn đề

Phương pháp bản đồ – biểu đồ là phương pháp nghiên cứu đặc trưng của Địa lý học. Đặc biệt, trong Địa lý kinh tế – xã hội, biểu đồ là công cụ trực quan để mô hình hóa các dữ liệu nghiên cứu và được sử dụng rất thường xuyên. Hiện nay, Địa lý học hiện đại đang chuyển mình mạnh mẽ với những ứng dụng công nghệ mới. Trong đó, SPSS là phần mềm phân tích dữ liệu đang được sử dụng nhiều trong các nghiên cứu của Địa lý kinh tế – xã hội. Ngoài những hỗ trợ về kỹ thuật phân tích, SPSS còn hỗ trợ xây dựng một số dạng biểu đồ. Trong nghiên cứu này, tác giả chỉ xin đề cập đến việc xây dựng và sử dụng biểu đồ hộp (Box plot) trong Địa lý kinh tế – xã hội.

2. Xây dựng biểu đồ hộp bằng phần mềm SPSS

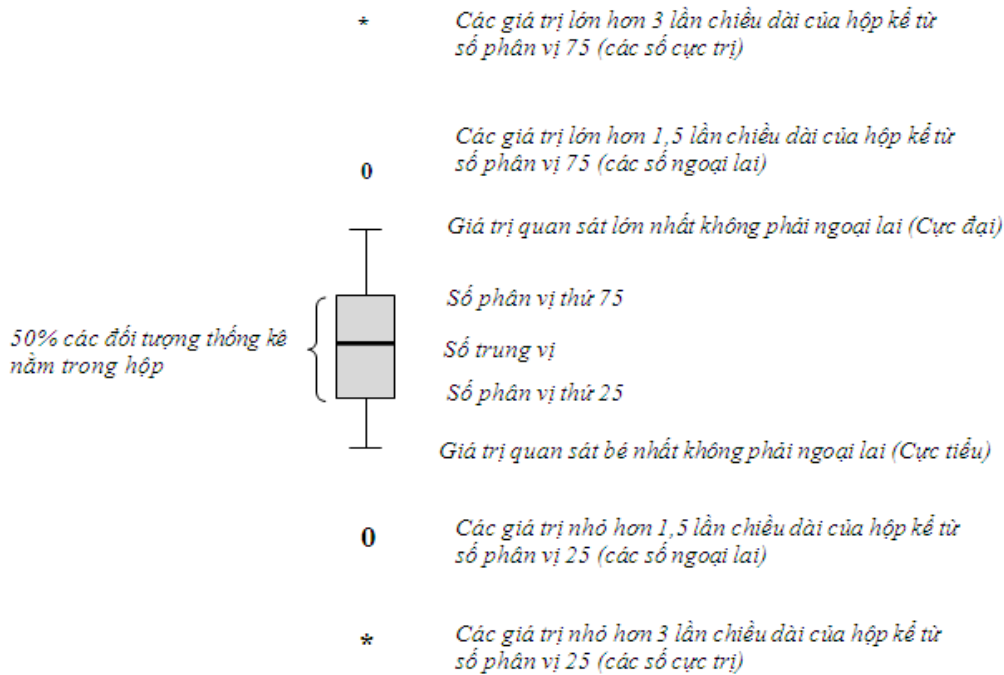
Biểu đồ hộp được phát minh bởi John Tukey năm 1977, đôi khi được gọi là biểu đồ hộp

và thanh (Box and Whisker plot). Bên cạnh việc cung cấp cô đọng hình ảnh về nơi mà các số liệu tập trung vào, cũng như hình dạng của từng phân phối, biểu đồ hộp còn giúp phát hiện các giá trị cực trị, ngoại lai, giá trị cực đại, cực tiểu...

Biểu đồ hộp được tạo ra khi sử dụng thủ tục Explore trong SPSS. Nó được vẽ dựa trên tứ phân vị (Tứ phân vị là đại lượng mô tả sự phân bố và sự phân tán của tập dữ liệu. Tứ phân vị có 3 giá trị, đó là tứ phân vị thứ nhất, thứ nhì, và thứ ba. Ba giá trị này chia một tập hợp dữ liệu (đã sắp xếp dữ liệu theo trật từ bé đến lớn) thành 4 phần có số lượng quan sát đều nhau). Đường nằm ngang trong hộp thể hiện giá trị trung vị của nhóm. Cạnh ngang (điểm chốt) dưới của hộp thể hiện phân vị thứ 25 và cạnh ngang trên thể hiện phân vị thứ 75. (Hiểu đơn giản: trung vị sẽ chia 1 nhóm các số liệu được sắp xếp theo thứ tự tăng dần thành 2 phần có số lượng đối tượng thống kê bằng nhau và từng cạnh

ngang của hộp một lần nữa chia tiếp từng phần nửa ra làm hai phần bằng nhau. Như vậy sẽ có 50% các đối tượng thống kê nằm trong hộp). Vạch ngang dưới thể hiện trị số quan sát nhỏ nhất mà không phải là số ngoại lai (outlier) và vạch ngang trên thể hiện trị số quan sát lớn nhất mà không phải là một số ngoại lai. Ký hiệu tròn trên biểu đồ

thể hiện các số ngoại lai (các trị số lớn hơn 1,5 lần chiều dài của hộp từ số phân vị thứ 75). Ký hiệu hoa thị trên biểu đồ thể hiện các số cực trị (các trị số lớn hơn 3 lần chiều dài của hộp kể từ số phân vị thứ 75). Chiều cao của hộp là khoảng biến thiên tứ phân vị (sự chênh lệch giá trị 2 cạnh nằm ngang).



Hình 1. Chú giải về một biểu đồ hộp

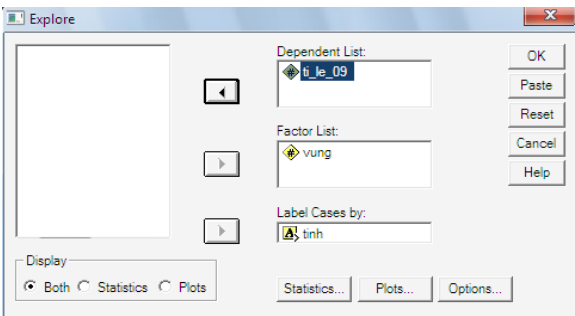
Giả sử có bảng số liệu dưới dạng Excel về tỉ lệ dân thành thị các tỉnh phân theo vùng của nước ta năm 2009 (Ti_le_09). Chúng ta tiến hành chuyển bảng số liệu từ Excel sang SPSS (có đuôi .sav): mở SPSS → Open file → trong ô Type chọn dạng .xls → chọn tên file Excel cần chuyển.

	A	B	C
1	Tỉnh	Vùng	Ti_le_09
2	Hà Nội	ĐB sông Hồng	41.0
3	Vĩnh Phúc	ĐB sông Hồng	22.5
4	Bắc Ninh	ĐB sông Hồng	23.5
5	Quảng Ninh	ĐB sông Hồng	51.9
6	Hải Dương	ĐB sông Hồng	19.0
7	Hải Phòng	ĐB sông Hồng	46.1
8	Hưng Yên	ĐB sông Hồng	12.1

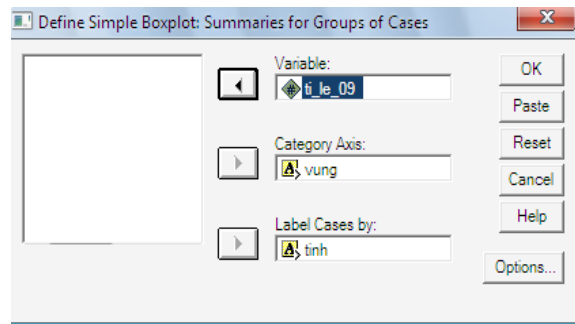
	tin	ving	ti_le_09
1	Hà Nội	ĐB sông Hồng	41.0
2	Vĩnh Phúc	ĐB sông Hồng	22.5
3	Bắc Ninh	ĐB sông Hồng	23.5
4	Quảng Ninh	ĐB sông Hồng	51.9
5	Hải Dương	ĐB sông Hồng	19.0
6	Hải Phòng	ĐB sông Hồng	46.1
7	Hưng Yên	ĐB sông Hồng	12.1
8	Thái Bình	ĐB sông Hồng	9.7
9	Hà Nam	ĐB sông Hồng	9.5
10	Nam Định	ĐB sông Hồng	17.6

Sau khi nguồn dữ liệu đưa vào SPSS được hiệu chỉnh, để tạo ra được biểu đồ hộp, có thể tiến hành theo hai cách:

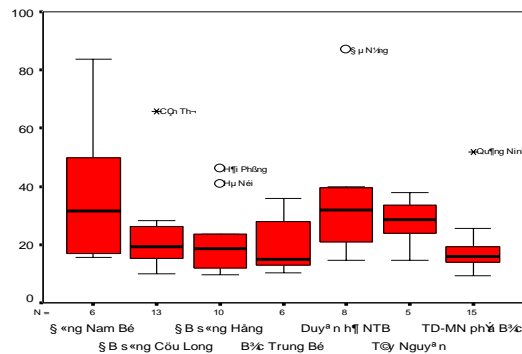
+ Cách 1: chọn lệnh Analyze → Descriptive Statistics → Explore. Ở thẻ Explore lần lượt đưa các biến vào các ô như hình dưới đây, các lệnh trong các thẻ phụ được mặc định, sau đó nhấn OK. Thực hiện theo cách này, ngoài biểu đồ hộp, SPSS còn tạo ra các bảng số liệu thống kê đi kèm như các giá trị trung bình, trung vị, cực đại, cực tiểu...



+ Cách 2: Tiến hành thực hiện lệnh: Graphs → Box plot → Simple → Define → Đưa các trường vào các ô theo như lệnh thực hiện trên hình vẽ → OK.



Kết quả được thể hiện trên một file Output mới. Tại đây, chúng ta có thể nhấp đúp chuột vào biểu đồ để tiến hành những hiệu chỉnh cần thiết về các định dạng của biểu đồ như màu sắc, font chữ... Sau khi hoàn thiện xong chúng ta có thể copy biểu đồ sang word. Ta được biểu đồ hộp có dạng như sau:



Hình 2. Biểu đồ hộp về tỉ lệ dân thành thị các vùng Việt Nam năm 2009

3. Ví dụ về sử dụng biểu đồ hộp trong Địa lí kinh tế – xã hội

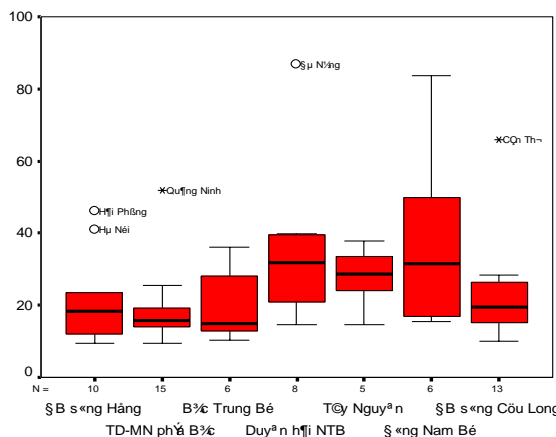
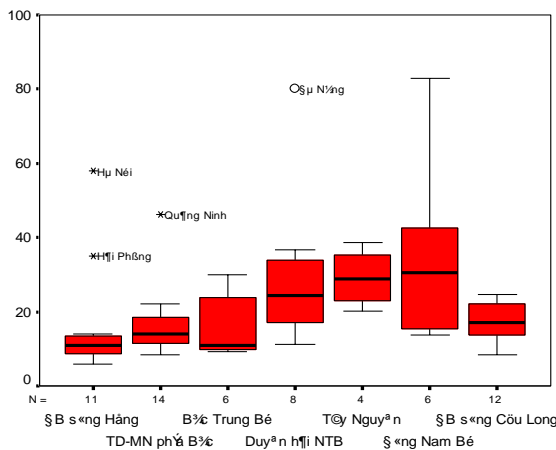
Biểu đồ hộp cho phép thể hiện rõ sự khác nhau về các chỉ tiêu kinh tế – xã hội của các đối tượng thống kê theo từng nhóm nước, khu vực hay vùng lãnh thổ. Giả sử để phân tích tình hình đô thị hóa của các vùng lãnh thổ nước ta qua hai năm 2000 và 2009. Biểu đồ hộp, được thành lập cho hai năm với các bước như trên, phản ánh rõ thực trạng phân hóa lãnh thổ và những chuyển biến trong bức tranh đô thị hóa các vùng ở nước ta.

Một số thống kê mô tả về tỉ lệ dân thành thị (%) ở các vùng kinh tế

Năm	Vùng kinh tế	Trung bình	Trung vị	Cực tiểu	Cực đại
2000	ĐB sông Hồng	16,7	10,8	5,8	57,9

	TD-MN phía Bắc	16,5	14,1	8,3	46,2
	Bắc Trung Bộ	15,7	10,8	9,2	29,9
	Duyên hải Nam Trung Bộ	30,3	24,4	11,2	80,1
	Tây Nguyên	29,1	28,8	20,2	38,6
	Đông Nam Bộ	35,9	30,5	13,8	82,9
	ĐB sông Cửu Long	17,3	17,2	8,5	24,6
2009	ĐB sông Hồng	21,9	18,5	9,5	46,1
	TD-MN phía Bắc	18,6	15,8	9,4	51,9
	Bắc Trung Bộ	19,6	15,0	10,4	36,0
	Duyên hải Nam Trung Bộ	35,8	31,9	14,6	86,9
	Tây Nguyên	27,7	28,6	14,7	37,9
	Đông Nam Bộ	38,2	31,6	15,6	83,7
	ĐB sông Cửu Long	22,9	19,4	10,0	65,9

Nhìn vào hai biểu đồ hộp và bảng số liệu về một số thống kê mô tả (sau khi đã tiến hành một số thủ tục xoay bảng trong SPSS) có thể thấy một số nét chính sau:



Hình 3. Biểu đồ hộp về tỉ lệ dân thành thị các vùng Việt Nam năm 2000 và 2009 (%)

- Với 11 tỉnh được thống kê năm 2000 và 10

tỉnh thống kê năm 2009, Đồng bằng sông Hồng (trừ Hà Nội và Hải Phòng) có tỷ lệ dân số đô thị khá thấp, tuy đang có xu hướng tăng lên. Trị số cực tiểu là thấp hơn cực tiểu của các vùng khác (năm 2000 là 5,8% và năm 2009 là 9,5%). Số trung vị cũng ở mức thấp, năm 2000 là 10,8% và năm 2009 là 18,5% (Số trung vị 18,5% năm 2009 có nghĩa là: 1/2 số tỉnh của đồng bằng có tỉ lệ đô thị hóa thấp hơn hoặc bằng 18,5% và 1/2 số tỉnh còn lại là bằng hoặc lớn hơn 18,5%). Khoảng biến thiên tứ phân vị (chiều cao của hộp) tương đối nhỏ, chỉ trên dưới 6%, chứng tỏ sự chênh lệch về tỉ lệ đô thị hóa của các tỉnh trong vùng không lớn. Phân phối năm 2000 khá cân đối nhưng đến năm 2009 hơi bị lệch trái (số trung vị nằm trên khoảng giữa của hộp), chứng tỏ tỉ lệ đô thị hóa các tỉnh đang tăng lên. Trên biểu đồ hộp của năm 2000, Hà Nội và Hải Phòng là các cực trị, nhưng trên biểu đồ năm 2009 chúng là các giá trị ngoại lai, điều này cho thấy rằng chính các thành phố lớn này làm thay đổi căn bản bức tranh đô thị hóa ở đồng bằng sông Hồng và khoảng cách về đô thị hóa giữa 2 thành phố lớn này với các địa phương khác đang dần thu hẹp. Trên thực tế, ngoài 2 đô thị lớn kể trên, có hai thành phố tương đối lớn là Hải Dương và Nam Định, ngoài ra là các thị xã nhỏ. Sự phát triển khá nhanh của các đô thị trung bình và nhỏ ở đồng bằng sông Hồng đã giúp làm thay đổi phần nào mẫu hình đô thị hóa của vùng.

- Trung du và miền núi phía Bắc cũng có tỉ lệ đô thị hóa thấp và có một cực trị là Quảng Ninh

(51,9% năm 2009). Các trung vị và các chốt (25% và 75%) năm 2000 ở vị trí cao hơn so với Đồng bằng sông Hồng, nhưng đến năm 2009, trung vị và chốt 75% đã ở mức thấp hơn, chứng tỏ tốc độ đô thị hóa ở Đồng bằng sông Hồng diễn ra nhanh hơn. Độ biến thiên có xu hướng thu hẹp lại và phân phối tương đối cân.

- Bắc Trung Bộ không có cực trị và ngoại lai, nhưng phân phối trải rộng và lệch phải mạnh hơn vì tỉ lệ dân đô thị là khá cao ở Thừa Thiên – Huế (36%) và Quảng Trị (28% năm 2009). Các tỉnh còn lại là Thanh Hóa, Nghệ An, Hà Tĩnh, Quảng Bình có tỉ lệ thấp hơn hẳn, thấp nhất là Thanh Hóa (10,4%).

- Duyên hải Nam Trung Bộ có bức tranh đô thị hóa tương đối sáng màu hơn so với Bắc Trung Bộ. Các giá trị trung vị (24,4% năm 2000 và 31,9% năm 2009) và các cực tiểu, cực đại đều nằm ở vị trí cao hơn, phân phối trải rộng hơn và tương đối cân. Đà Nẵng là một ngoại lai. Tỉ lệ dân thành thị ở Đà Nẵng đạt 86,9% năm 2009.

- Tây Nguyên có tỉ lệ dân thành thị tương đối cao so với Đồng bằng sông Hồng, Trung du và miền núi phía Bắc và Bắc Trung Bộ. Các giá trị trung vị, cực đại và cực tiểu đều nằm ở vị trí cao hơn. Khoảng biến thiên không quá lớn và phân phối tương đối cân. Trên thực tế, các tỉnh ở Tây Nguyên tuy có số lượng dân thành thị nhỏ, nhưng do qui mô dân số nhỏ và dân cư đa phần tập trung trong các đô thị nên tỉ lệ dân thành thị đạt khá cao.

- Đông Nam Bộ có tỉ lệ đô thị hóa cao nhất cả nước. Trung vị, cực tiểu và cực đại của vùng này đều nằm ở mức cao hơn so với các vùng khác. Phân phối khá cân bằng. Khoảng biến thiên từ phân vị khá lớn là do tỉ lệ dân thành thị cao hơn

hẳn ở các tỉnh Đồng Nai, Bà Rịa – Vũng Tàu và đặc biệt là ở Tp.Hồ Chí Minh. Khoảng phân vị từ chốt 75% đến giá trị cực đại bị kéo dài ra, khiến ta có cảm giác như Tp.Hồ Chí Minh đang cố tách ra trở thành một ngoại lai, nhưng kì thực là nó đang bị thu về. Nếu xét các năm trước 2000 thì rõ ràng Tp.Hồ Chí Minh là một ngoại lai và chi phối gần như toàn bộ bức tranh đô thị hóa của Đông Nam Bộ. Nhưng với sự phát triển năng động của vùng kinh tế trọng điểm phía Nam, tốc độ đô thị hóa của vùng tăng lên, nhất là ở Bình Dương, Đồng Nai, Bà Rịa – Vũng Tàu và Tây Ninh. Vì vậy, đến năm 2000 Tp.Hồ Chí Minh không còn là một ngoại lai nữa.

- Đồng bằng sông Cửu Long nhìn chung tỉ lệ đô thị hóa còn khá thấp. Năm 2000, phân phối khá cân đối và khoảng biến thiên từ phân vị nhỏ, thì đến 2009 phân phối có xu hướng lệch phải và khoảng biến thiên được mở rộng ra hơn. Cần Thơ, sau khi tách ra khỏi Hậu Giang và trở thành thành phố trực thuộc Trung ương, đã tạo thành một cực trị trong bức tranh đô thị hóa của vùng. Tỉ lệ đô thị hóa của Cần Thơ vượt hơn hẳn các tỉnh khác, đạt mức 65,9% năm 2009.

4. Kết luận

Biểu đồ là công cụ không chỉ giúp phản ánh một cách trực quan các số liệu thống kê, mà thông qua đó còn có thể phát hiện nhiều vấn đề mới. Sử dụng và phổ biến các tính năng của SPSS trong dạy học và nghiên cứu địa lí là rất quan trọng, trong đó có việc sử dụng các dạng biểu đồ. Tác giả đã tập trung trình bày khá chi tiết về cách thức xây dựng và ý nghĩa biểu hiện của dạng biểu đồ hộp, đồng thời đưa ra ví dụ minh họa về việc vận dụng cụ thể trong phân tích một vấn đề của Địa lí kinh tế – xã hội.

TÀI LIỆU THAM KHẢO

- [1] Nguyễn Văn Liệu và nnk (2000), *SPSS - Ứng dụng phân tích dữ liệu trong quản trị kinh doanh và khoa học tự nhiên – xã hội*, NXB Giao thông vận tải.
- [2] Đỗ Thị Minh Đức (2006), “Đô thị hóa Việt Nam trong bối cảnh thế giới đô thị hóa”, *Tạp chí Khoa học DHSP Hà Nội*.
- [3] Hoàng Trọng, Chu Nguyễn Mộng Ngọc (2005), *Phân tích dữ liệu nghiên cứu với SPSS*, NXB Thống kê.
- [4] Tổng cục thống kê (2010), *Niên giám thống kê 2009*, NXB Thống kê.

