

GIỚI THIỆU CÁC LOẠI ĐỒ THỊ THƯỜNG ÁP DỤNG TRONG THỐNG KÊ MÔ TẢ DOMINIQUE HAUGHTON

Nguyễn Phong

Vụ trưởng vụ Xã hội Môi trường- TCTK

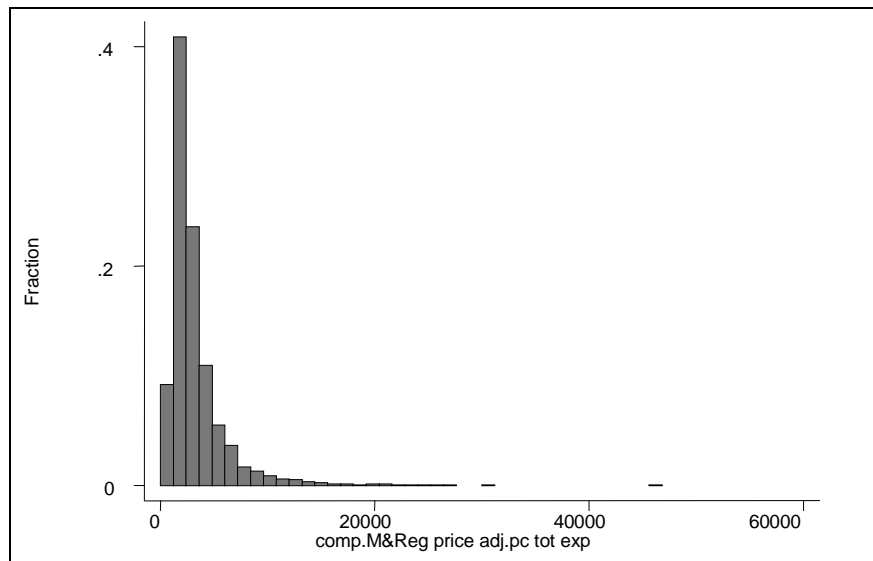
1. Đồ thị cột, đồ thị hình hộp, độ lệch của số liệu

Các đồ thị hình cột (xem Hình 1a và 1b) về chi tiêu thực tế bình quân đầu người của 5.999 hộ gia đình Việt nam bao gồm hoặc không bao gồm các giá trị ngoại lai đều cho thấy mức độ lệch phải giống nhau. Điều này có nghĩa là chi tiêu thực tế bình quân đầu người có phân bố không đối xứng. Nói cách khác, số người có chi tiêu thực tế bình quân đầu người ở mức thấp lớn hơn số người có chi tiêu thực tế bình quân ở mức cao. Vì phân bố lệch phải nên giá trị trung bình của chi tiêu thực tế bình quân đầu người lớn hơn giá trị trung vị (3189 ngàn đồng so với 2397 ngàn đồng, đã loại các giá trị ngoại lai).

Đồ thị trong Hình 1a có 50 cột. Lệnh "sum rlpdex1,detail" của phần mềm Stata cho thấy giá trị nhỏ nhất

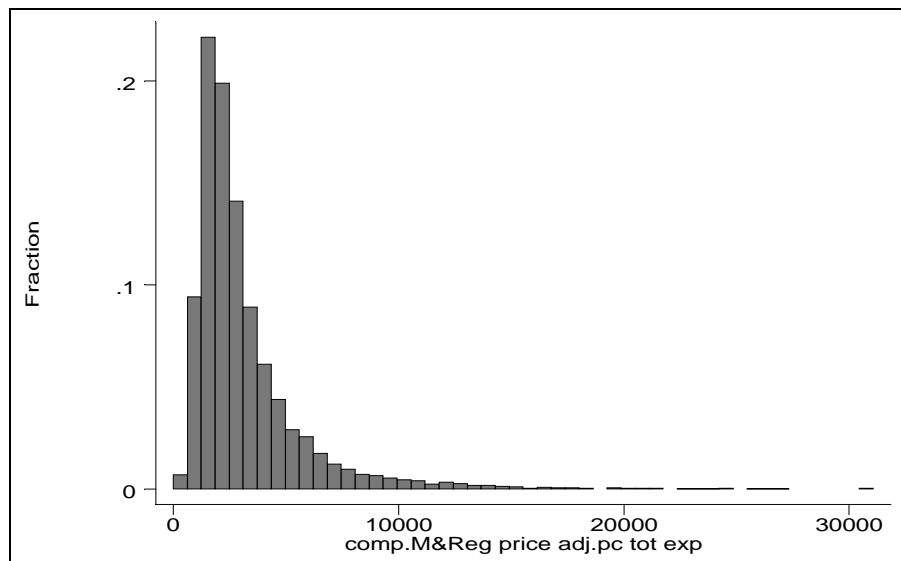
của chi tiêu thực tế bình quân là 357,318 ngàn đồng và giá trị lớn nhất là 45.801,71 ngàn đồng. Chiều rộng của mỗi cột là 908,89 được tính bằng cách lấy giá trị lớn nhất trừ đi giá trị nhỏ nhất và chia kết quả tìm được cho 50. Vì vậy, cột thứ nhất trong đồ thị sẽ có chiều rộng là 908,89, là khoảng cách từ giá trị 357,32 đến 1266,21. Điều này có nghĩa là gần 10% số hộ trong tổng số 5.999 hộ có mức chi tiêu thực tế bình quân đầu người một năm nằm trong khoảng từ 357,32 ngàn đồng đến 1266,21 ngàn đồng. Tương tự, cột thứ 2 của đồ thị cũng có chiều rộng là 908,89 nhưng nằm trong khoảng từ 1.266,21 và 2.175,10. Từ đồ thị này chúng ta có thể thấy gần 40% trong tổng số 5.999 hộ có mức chi tiêu thực tế bình quân đầu người một năm nằm trong khoảng từ 1.266,21 đến 2.175,10 ngàn đồng.

HÌNH 1A: ĐỒ THỊ HÌNH CỘT CỦA CHI TIÊU THỰC TẾ BÌNH QUÂN ĐẦU NGƯỜI MỘT NĂM CỦA 5.999 HỘ GIA ĐÌNH VIỆT NAM



Cách lý giải đồ thị trong Hình 1b cũng tương tự như trên, trừ một điểm là ở đồ thị này đã loại bỏ các giá trị ngoại lai.

HÌNH 1B: ĐỒ THỊ HÌNH CỘT CỦA CHI TIÊU THỰC TẾ BÌNH QUÂN ĐẦU NGƯỜI CỦA 5.999 HỘ GIA ĐÌNH VIỆT NAM (ĐÃ LOẠI BỎ CÁC GIÁ TRỊ NGOẠI LAI).



Đồ thị hình hộp là một công cụ đồ họa rất hữu ích giúp nhìn thấy ngay được vị trí, mức độ dàn trải, độ lệch và các giá trị ngoại lai của một bộ số liệu. Cạnh đáy dưới của hình hộp (xem Hình 2) là tứ phân

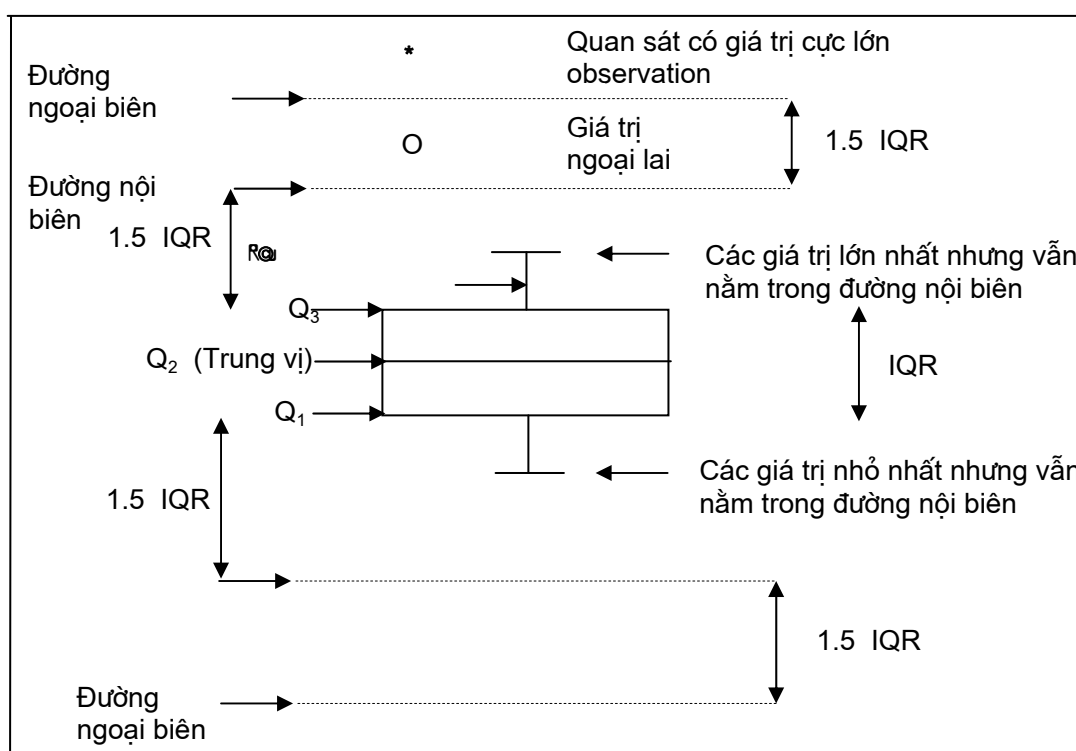
vị thứ nhất; 25% các giá trị chi tiêu thực tế bình quân đầu người một năm nằm dưới tứ phân vị thứ nhất này và 75% các giá trị chi tiêu thực tế bình quân đầu người một năm nằm trên mức này. Cạnh đáy trên

của hình hộp là tứ phân vị thứ ba; 75% các giá trị chi tiêu thực tế bình quân đầu người một năm nằm dưới tứ phân vị thứ ba này và 25% các giá trị chi tiêu thực tế bình quân đầu người một năm nằm trên mức này. Đoạn thẳng nằm trong hình hộp và song song với 2 cạnh đáy là giá trị trung vị của chi tiêu thực tế bình quân đầu người một năm và là tứ phân vị thứ hai. Chiều cao

của hình hộp là hiệu số giữa tứ phân vị thứ ba và thứ nhất và được gọi là miền trong tứ phân vị (interquartile range - IQR). Cộng 1,5 giá trị của IQR xuống phía dưới đối với tứ phân vị thứ nhất và lên phía trên

với tứ phân vị thứ ba ta sẽ có hai đường nội biên. Hai đoạn thẳng nằm phía trên và phía dưới hình hộp lần lượt biểu thị các giá trị lớn nhất và nhỏ nhất nhưng vẫn nằm trong các đường nội biên. Hai đoạn thẳng này được gọi là râu (whiskers). Tương tự như vậy, hai đường ngoại biên được thiết lập bằng cách cộng thêm 3 lần giá trị IQR. Các giá trị nằm ngoài các đường ngoại biên là các giá trị cực nhỏ hoặc cực lớn, còn các giá trị nằm giữa các đường nội biên và ngoại biên là các giá trị ngoại lai. Phần mềm này cho chúng ta biết hai râu nhưng không hiển thị các đường biên (nội biên và ngoại biên).

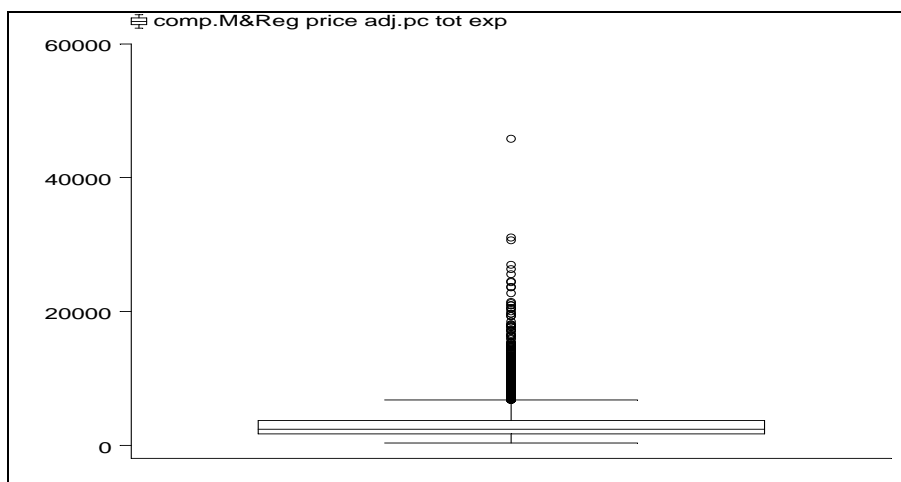
HÌNH 2: CÁC ĐẶC TRƯNG CỦA ĐỒ THỊ HÌNH HỘP



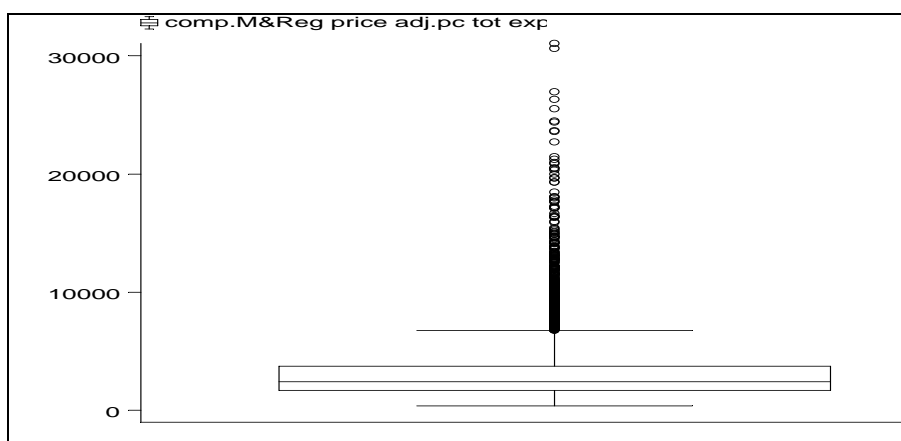
Hình 3a và 3b là các đồ thị hình hộp của chi tiêu thực tế bình quân đầu người của bộ số liệu VLSS 1998, bao gồm và không bao gồm các quan sát ngoại lai. Khoảng cách từ đường trung vị đến cạnh

đáy dưới gần hơn đến cạnh đáy trên của hình hộp ở các đồ thị này cho thấy đặc trưng tương tự của số liệu như ở đồ thị hình cột, tức là biến chi tiêu thực tế bình quân đầu người có phân bố lệch phải.

HÌNH 3A: ĐỒ THỊ HÌNH HỘP CHI TIÊU THỰC BÌNH QUÂN ĐẦU NGƯỜI CỦA 5.999 HỘ GIA ĐÌNH VIỆT NAM



HÌNH 3B: ĐỒ THỊ HÌNH HỘP CHI TIÊU THỰC TẾ BÌNH QUÂN ĐẦU NGƯỜI CỦA 5.999 HỘ GIA ĐÌNH VIỆT NAM (ĐÃ LOẠI BỎ CÁC GIÁ TRỊ NGOẠI LAI)



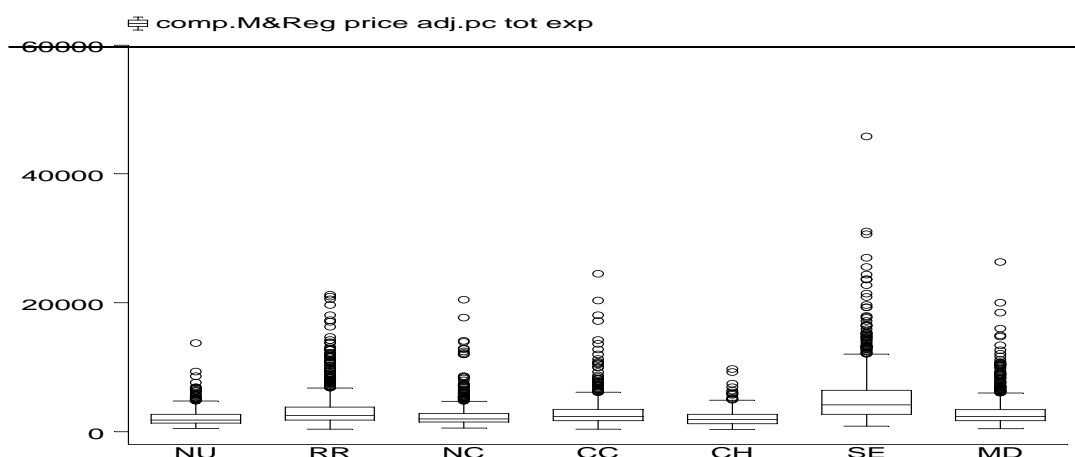
Trong các Hình 1a và 3a có một giá trị ngoại lai với chi tiêu thực tế bình quân đầu người cực lớn. Đó là 1 hộ phi nông nghiệp ở khu vực thành thị thuộc vùng Đông Nam bộ (bao gồm thành phố Hồ Chí Minh). Chi tiêu thực tế bình quân đầu người của hộ này lên đến gần 46 triệu đồng 1 năm. Chủ hộ là nam giới với trình độ giáo dục đại học và là chủ 1 doanh nghiệp tư nhân. Số năm đi học bình quân

của những người trưởng thành trong hộ này rất cao (17 năm). Điều này chứng tỏ vợ của chủ hộ này cũng có trình độ giáo dục rất cao. Mặt khác, qui mô của hộ này lại rất nhỏ, chỉ có 2 người và không có con.

Đồ thị hình hộp đặc biệt có ích khi so sánh các nhóm con. Hình 4 đưa ra đồ thị hình hộp cho mỗi vùng của Việt nam, theo thứ tự từ trái sang phải: Miền núi và trung du phía Bắc, Đồng bằng sông Hồng, Bắc

Trung bộ, Duyên hải miền Trung, Tây Nguyên, Đông Nam bộ và Đồng bằng sông Cửu Long. Cùng với các đặc trưng khác, phân bố chi tiêu thực tế bình quân đầu người của vùng Đông Nam bộ (trong đó có thành phố Hồ Chí Minh) rõ ràng cao hơn và có mức độ biến thiên lớn hơn (hình hộp cao hơn) so với các vùng còn lại.

HÌNH 4: CÁC ĐỒ THỊ HÌNH HỘP CHI TIÊU THỰC TẾ BÌNH QUÂN ĐẦU NGƯỜI CỦA 5.999 HỘ GIA ĐÌNH VIỆT NAM CHIA THEO VÙNG



(còn nữa)

Tài liệu tham khảo

Chiu, S.T. "Lựa chọn độ rộng cho ước lượng mật độ Kernel", *Biên niên sử thống kê*, 19, 1883-1905 (1991).

Hintze, J. and R.D. Nelson. "Đồ thị viôlong: Sự kết hợp giữa đồ thị hình hộp và đồ thị mật độ", *Tạp chí nhà thống kê Mỹ*, 52, 181-184 (1998).

Tổng cục Thống kê. *Niên giám thống kê*. Hà nội: Nhà xuất bản thống kê (2001).

Rousseeuw, P., I. Ruts and J. Tukey. "Đồ thị hình túi: Đồ thị hình hộp biến thiên hai chiều", *Tạp chí nhà thống kê Mỹ*, 53, 382-387 (1999).

Scott, D.W. "Về đồ thị hình cột tối ưu và đồ thị hình cột dựa trên số liệu thực tế", *Biometrika*, 66, 605-610 (1979).

Silverman, B. *Ước lượng mật độ cho số liệu thống kê và phân tích số liệu*. London: Chapman and Hall (1986).

Công ty Stata. "Lệnh kdensity", *Sách hướng dẫn Stata cho phiên bản 6*, tập 2, 144-151 (1999).

Tukey, J. *Phân tích số liệu thăm dò*. Reading, MA: Addison-Wesley (1977).

Wand, M.P. "Lựa chọn số cột cho đồ thị hình cột dựa trên các bộ số liệu thực tế", *Tạp chí nhà Thống kê*, 51, 59-64 (1997).