



PHÂN TÍCH HỒI QUY XU THẾ VÀ MỘT ỨNG DỤNG THỰC VỊ

Trần Văn Lý, Lê Thị Hải Yên, Nguyễn Huyền Trang, Trần Kim Yến, Bùi Minh Trung và Lâm Quốc Toàn

Khoa Khoa học Tự nhiên, Trường Đại học Cần Thơ

Thông tin chung:

Ngày nhận: 07/03/2016

Ngày chấp nhận: 29/08/2016

Title:

The trend regression analysis and an interesting application

Từ khóa:

Hồi qui, chuỗi thời gian, khuynh hướng, mờ, tương quan, dự báo

Keywords:

Regression, time series, trend, fuzzy, correlation, forecast

ABSTRACT

The article introduces a method of statistical analysis, which is applied appropriately for data having cyclical and “seasonal” characteristics. The data will be modeled in the form of the trend regression model, which is established from the time variable and “season pseudo” variables. The method is illustrated interestingly on the data collected at the canteen of Can Tho University. We also test this data for the ARIMA model, a model in time series analysis.

TÓM TẮT

Bài viết giới thiệu một phương pháp phân tích thống kê được áp dụng phù hợp cho các dữ liệu có tính tuần hoàn, “mùa vụ”. Dữ liệu sẽ được mô hình hoá theo dạng mô hình hồi quy xu thế, được thiết lập dựa trên biến thời gian và các biến “giả mùa”. Phương pháp được áp dụng minh họa rất thú vị trên bộ dữ liệu kinh doanh của canteen Văn phòng Đoàn Trường Đại học Cần Thơ. Chúng tôi cũng thử áp dụng bộ dữ liệu này với mô hình ARIMA, một mô hình rất hay dùng trong phân tích chuỗi thời gian.

Trích dẫn: Trần Văn Lý, Lê Thị Hải Yên, Nguyễn Huyền Trang, Trần Kim Yến, Bùi Minh Trung và Lâm Quốc Toàn, 2016. Phân tích hồi quy xu thế và một ứng dụng thực vị. Tạp chí Khoa học Trường Đại học Cần Thơ. 45a: 118-125.

1 GIỚI THIỆU

Ứng dụng phân tích hồi quy được giới thiệu lần đầu tiên từ năm 1877 bởi Sir Francis Galton (Aczeta and Sounderpandian, 2009; Cleff, 2011). Một trong những mục tiêu quan trọng của phân tích hồi quy là đưa ra khả năng dự báo cho biến phụ thuộc dựa trên dữ liệu thu thập của một số biến độc lập khác. Trong trường hợp dữ liệu được thu thập theo một biến đổi thời gian (có biến độc lập là biến thời gian) thì mô hình hồi quy được thiết lập là một mô hình hồi quy xu thế. Theo từng khoảng thời gian cục bộ, mô hình có thể có xu thế tăng hay giảm thuần túy, hoặc hỗn hợp biến thiên ngẫu nhiên hay biến thiên tuần hoàn với tính chu kỳ, “mùa vụ”. Trong trường hợp này, việc vận dụng chặt chẽ lý thuyết hồi quy cùng với việc mô hình hóa phù hợp cho bộ dữ liệu (với sự hỗ trợ của các biến “giả

mùa”) sẽ giúp chúng ta thu được những kết quả dự báo phù hợp hơn. Xem xét những nền tảng chặt chẽ của lý thuyết hồi quy để xây dựng một mô hình như vậy, thường được gọi là “Mô hình phân tích hồi quy xu thế”, là phần nội dung thứ nhất được trình bày trong bài báo này.

Một mảng rất gần với mô hình phân tích hồi quy xu thế là các dạng ứng dụng mô hình hoá trong lý thuyết phân tích chuỗi thời gian (Groebner *et al.*, 2011; Trần Hùng Thao, 2013), như mô hình trung bình trượt MA (Moving Average), mô hình tự hồi quy AR (AutoRegressive) hay các dạng kết hợp như ARMA (AutoRegressive Moving Average), ARIMA (AutoRegressive Integrated Moving Average). Trong mỗi tình huống ứng dụng cụ thể, vấn đề đặt ra là vận dụng một mô hình phân tích chuỗi thời gian hay mô hình phân tích hồi quy xu

thế là phù hợp? Xem xét cho vấn đề này, phần nội dung thứ hai của bài báo sẽ đưa ra các áp dụng thử nghiệm trên một bộ dữ liệu cụ thể. Chúng tôi sẽ vận dụng thử nghiệm và quan sát kết quả dự báo trên cùng một bộ dữ liệu cho mô hình phân tích hồi quy xu thế và mô hình ARIMA, một mô hình khá phổ biến trong phân tích chuỗi thời gian.

Bộ dữ liệu sử dụng là các dữ liệu kinh doanh của căn tin Văn phòng Đoàn Trường Đại học Cần Thơ, được lấy từ ngày 01/02/2015 đến 30/09/2015, rơi vào 3 học kỳ (học kỳ 2, học kỳ hè năm học 2014-2015 và học kỳ 1 năm học 2015-2016). Bộ dữ liệu này phù hợp với điều kiện nghiên cứu của chúng tôi và cũng đáp ứng được mục đích phân tích của bài báo. Đây là bộ dữ liệu có tính “mùa vụ”, các dữ liệu thay đổi có tính chu kỳ theo thời gian, phụ thuộc vào lịch học của các ngày trong tuần và của các học kì trong năm. Các xử lý dữ liệu thống kê được chúng tôi thực hiện trên phần mềm R.

Việc giới thiệu chi tiết về dữ liệu cũng như về các vấn đề nội dung đã nêu sẽ được lần lượt trình bày trong ba mục còn lại của bài báo. Trong mục thứ 2 chúng tôi sẽ trình bày các nền tảng lý thuyết liên quan đến phương pháp phân tích hồi quy xu thế và mô hình ARIMA. Việc áp dụng hai mô hình này trên dữ liệu thực trên sẽ được giới thiệu ở mục 3. Cuối cùng, trong phần kết luận ở mục 4 chúng tôi đưa ra nhận xét về sự vận dụng phù hợp của hai dạng mô hình này, đồng thời nêu một số kết quả phân tích thống kê liên quan đến bộ dữ liệu.

2 CƠ SỞ LÝ THUYẾT

2.1 Xây dựng phương trình hồi quy xu thế

Phương pháp hồi quy xu thế được đề cập là phân tích hàm hồi quy giữa biến phụ thuộc x và biến thời gian t , $x = f(t)$. Nếu $f(t)$ là một hàm tuyến tính thì được gọi là có xu thế biến đổi tuyến tính. Trong những trường hợp khác ta sẽ gọi nó là xu thế không tuyến tính.

Để nghiên cứu xu thế biến đổi tuyến tính ta thành lập phương trình hồi quy

$$x(t) = at + b, \tag{1}$$

trong đó a và b là các hệ số hồi quy được xác định bởi:

$$a = \frac{\sum_{t=1}^n (x_t - \bar{x})(t - \bar{t})}{\sqrt{\sum_{t=1}^n (x_t - \bar{x})^2 (t - \bar{t})^2}}, \tag{2}$$

$$b = \bar{x} - a\bar{t}, \tag{3}$$

$$\bar{x} = \frac{1}{n} \sum_{t=1}^n x_t, \quad \bar{t} = \frac{1}{n} \sum_{t=1}^n t,$$

với \bar{x} , \bar{t} lần lượt là trung bình mẫu của biến x và t .

Chúng ta có thể nhận biết được xu thế biến đổi của dữ liệu x thông qua phân tích hệ số góc a . Dấu của hệ số a xác định xu thế tăng (khi $a > 0$) hoặc giảm (khi $a < 0$), còn trị tuyệt đối của a cho biết mức độ tăng giảm của dữ liệu theo biến đổi của thời gian t .

Chúng ta có thể khảo sát dữ liệu trong các khoảng thời gian khác nhau để phân tích xu thế. Khi đó căn cứ vào các hệ số góc a chúng ta có thể phân tích xu thế của dữ liệu qua các thời kỳ khác nhau.

2.2 Mô hình hồi quy với biến giả

Biến giả là biến được dùng để lượng hóa các biến chất lượng. Biến giả nhận hai giá trị bằng 1 hoặc bằng 0 (biến nhị phân). Với biến chất lượng có nhiều hơn hai phạm trù, ta dùng nhiều biến giả để lượng hóa.

Một số lưu ý:

i) Biến chất lượng có m phạm trù thì sẽ sử dụng $m - 1$ biến giả.

ii) Phạm trù được gán giá trị 0 là phạm trù cơ sở.

iii) Các hệ số góc đại diện cho chênh lệch giữa giá trị trung bình của các phạm trù với phạm trù cơ sở.

iv) Phạm trù có biến giả đặt cho lấy giá trị 1 là phạm trù so sánh.

2.3 Chuỗi thời gian

Phân tích chuỗi thời gian là dự báo giá trị của biến cần dự báo vào thời điểm hiện tại dựa trên cơ sở các giá trị của nó trong quá khứ cộng với một phần sai số, phần sai số này biến động ngẫu nhiên. Chuỗi dữ liệu phân tích phải có tính dừng, nếu không dừng ta tiến hành lấy sai phân đến khi chuỗi dừng.

2.3.1 Mô hình trung bình trượt tự hồi quy ARIMA(p,d,q)

Phương trình khái quát của mô hình ARIMA(p,d,q) có dạng như sau (Anderson *et al.*, 2011; Concetta and David, 2011):

$$Y = \delta + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q} + e_t, \tag{4}$$

trong đó: $\alpha_1, \dots, \alpha_p$ là các tham số tự hồi quy; β_1, \dots, β_q là các tham số trung bình trượt; μ là giá trị trung bình của chuỗi thời gian; $\delta = \mu(\beta_1 + \beta_2 + \dots + \beta_q)$ và e_t là nhiễu trắng.

Chúng tôi sử dụng phương pháp Box-Jenkin kết hợp với phần mềm R để xác định mô hình ARIMA cho lượng ăn vặt và giải khát bán được trong một ngày.

2.3.2 Phương pháp Box-Jenkins

Mục tiêu là tìm trong số tất cả các dạng của mô hình ARIMA một mô hình thích hợp nhất với bộ số liệu của hiện tượng đang nghiên cứu.

Bước 1: Nhận dạng mô hình.

Xác định p, q của mô hình ARIMA nhờ vào đồ thị tự tương quan ACF và tương quan riêng PACF.

Bước 2: Ước lượng các hệ số của mô hình.

Các hệ số của mô hình được xác định theo phương pháp bình phương tối thiểu (OLS_Ordinary Least Square) sao cho:

$$\sum(X_t - \hat{X}_t) \rightarrow \min. \tag{5}$$

Bước 3: Kiểm định mô hình.

Các hệ số của mô hình phải khác 0 (kiểm định t). Nếu có một hoặc nhiều hệ số không thỏa mãn ta sẽ loại bỏ ra mô hình AR hay MA đang xét.

Phần dư của mô hình là sai số giữa kết quả dự báo và số liệu thực tế ($\epsilon_t = X_t - \hat{X}_t$).

Bước 4: Dự báo.

Khi mô hình đã được kiểm định ta có thể dự báo cho tương lai.

3 ỨNG DỤNG PHƯƠNG PHÁP TRÊN DỮ LIỆU THỰC

Do mục đích của nhóm chúng tôi là tập trung phân tích về tác động của các yếu tố thời gian vào nhiệt độ và xu hướng bán ra của hai biến Giải Khát và An Vạt. Đồng thời xem xét tác động của hai yếu tố này vào doanh thu của căn tin, từ đó rút ra chiến lược để hoạt động kinh doanh của căn tin đạt hiệu quả tốt hơn. Bên cạnh đó, chúng tôi cũng sẽ đưa ra dự báo về tình hình kinh doanh của căn tin trong những ngày tiếp theo.

3.1 Giới thiệu bộ số liệu

Bộ dữ liệu bao gồm 37 biến:

+ Nhóm các biến thời gian: Ngày (t=1, 2, 3, 4, 5, ..., 242), Thu, Hocky.

+ Nhóm các biến giải khát (Giải Khát): Trà sữa, C2, cà phê, cà phê sữa, chanh leo, lipton, nước mía, nước suối, phô mai que, rau câu, trà đào, tắc xi

muội, trà bạc hà, trà chanh, trà đường, trà dâu, trà kiwi mà căn tin bán được mỗi ngày.

+ Nhóm các biến ăn vặt (An Vạt): Bánh flan, bánh khoai môn, bò viên, cá viên, xúc xích, dưa lưới, dưa tắc, há cảo, hải sản rau củ, khoai tây lắc.

+ Tổng doanh thu hàng ngày (Tongdoanhthu).

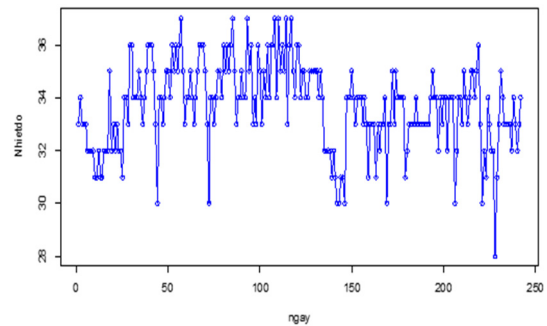
+ Nhiệt độ cao nhất mỗi ngày ở Cần Thơ (Nhiệt Do).

Trong đó, các loại nước uống được tính bằng ly, ngoại trừ C2 và nước suối tính bằng chai, caramem trung, bánh khoai môn, xúc xích, phô mai que được tính bằng cái, bánh flan, bò viên, cá viên, há cảo, hải sản rau củ, khoai tây lắc, rau câu được tính bằng đĩa (hoặc bát), doanh thu tính bằng đơn vị VNĐ, nhiệt độ được tính bằng độ C. Trong bài này chúng tôi phân tích thống kê sự phụ thuộc giữa các biến An Vạt, Giải Khát, Nhiệt Do, Tongdoanhthu, lượng bán được chung của An Vạt và Giải Khát (AV&GK) và các biến thời gian.

3.2 Thống kê mô tả

3.2.1 Nhiệt độ ở Cần Thơ

Dữ liệu khảo sát cho thấy rằng, tại thành phố Cần Thơ trong thời gian này có trung bình là 33.678°C; có xu hướng giảm và dần mát dịu trong khoảng thời gian về sau. Với điều kiện nhiệt độ như thế thì tình hình kinh doanh của căn tin có bị ảnh hưởng hay không? Câu hỏi này sẽ được chúng tôi làm rõ ở phần sau.



Hình 1: Biểu đồ nhiệt độ ở Cần Thơ

3.2.2 Thống kê mô tả cho dữ liệu chuỗi thời gian

+ So sánh đối chiếu sự khác nhau giữa giải khát và ăn vặt hàng ngày khi không quan tâm đến thời gian qua một số tham số đặc trưng được liệt kê trong Bảng 1:

Bảng 1: Một số tham số đặc trưng của các biến Giai Khat và An Vat

Biến	Min	Median	Mean	Max	SD	SE	NA's
Giai Khat	1	26	28.61	72	16.62	3.24	76
An Vat	0	28.5	30.64	96	17.22	3.76	76

Nhìn vào các giá trị mean, median, max của hai bảng thống kê mô tả của hai biến ta thấy sự chênh lệch khá lớn về số lượng bán ra của giải khát và ăn vặt. Trung bình lượng giải khát bán ra là ≈ 29 trong

khí lượng ăn vặt đạt ≈ 31 .

+ Bảng 2 và Bảng 3 liệt kê các tham số đặc trưng của các biến Giai Khat và An Vat vào các ngày khác nhau trong tuần:

Bảng 2: Một số tham số đặc trưng của biến Giai Khat theo tuần

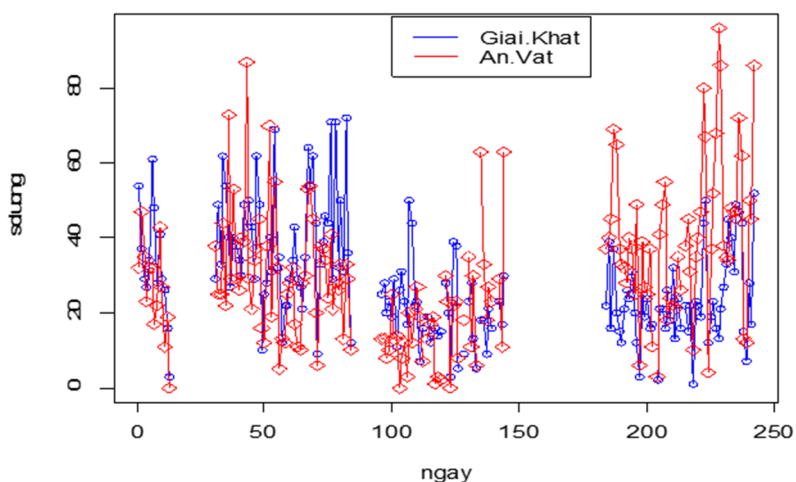
	N	NaN	Min	Median	Mean	Max	VAR	SD	SE
Chunhat	19	16	1	19	23.74	71	369.09	19.21	4.41
Thuhai	25	10	9	24	27.36	50	97.32	9.87	1.97
Thuba	25	10	12	29	28.12	50	113.03	10.63	2.13
Thutu	25	10	3	26	27.8	64	223.5	14.95	2.99
Thunam	25	9	16	33	34.96	72	252.62	15.89	3.18
Thusau	25	9	3	30	32.28	71	355.88	18.86	3.77
Thubay	22	12	5	21	24.32	54	213.94	14.63	3.12

Bảng 3: Một số tham số đặc trưng của biến An Vat theo tuần

	N	NaN	Min	Median	Mean	Max	VAR	SD	SE
Chunhat	19	16	3	13	23.79	87	518.84	22.78	5.23
Thuhai	25	10	12	35	34.28	63	151.46	12.31	2.46
Thuba	25	10	11	33	33.92	70	239.83	15.49	3.1
Thutu	25	10	0	29	35.12	96	514.69	22.69	4.54
Thunam	25	9	0	32	35	86	522.5	22.86	4.57
Thusau	25	9	0	30	31.24	67	331.52	18.21	3.64
Thubay	22	12	2	14.5	17.95	49	168.14	12.97	2.76

Khi xem xét dữ liệu theo các ngày trong tuần, nhìn chung ta thấy rằng lượng bán ra của Giai Khat và An Vat khá cao ở giữa tuần, giảm mạnh vào cuối tuần do có quá ít lớp học. Điều này cần được phân tích thống kê cụ thể hơn để giúp cho người

quản lý căn tin cân đối trong việc bố trí nhân viên làm việc cũng như lượng hàng cần cung cấp vào các ngày trong tuần sao cho phù hợp nhất, để căn tin hoạt động đạt hiệu quả cao nhất.



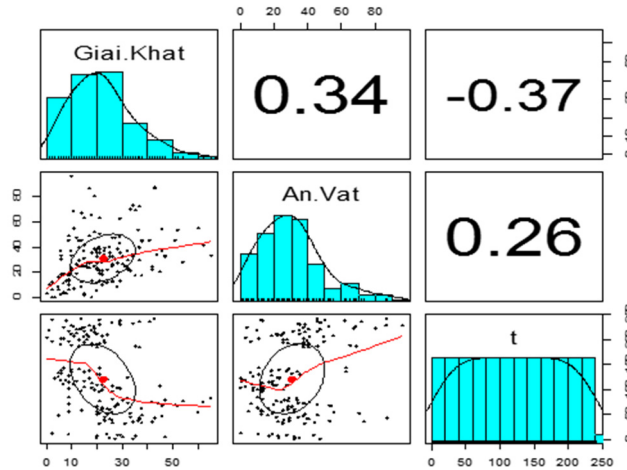
Hình 2: Biểu đồ chuỗi thời gian của biến Giai Khat và An Vat

Biểu đồ trên cho thấy số lượng ăn vặt và giải khát bán ra có sự chênh lệch và biến động không đều. Ta thấy rằng số lượng của cả ăn vặt và giải khát bán ra chênh lệch không đáng kể trong tháng hai và tháng tư và số lượng này được bán ra nhiều hơn so với giải khát vào các tháng còn lại (trừ tháng năm).

Nếu người quản lý không phân tích xu thế trong

số liệu, họ có thể kết luận họ cần phải có số lượng tương tự của cả ăn vặt và giải khát vì vậy có thể dẫn đến thừa giải khát nhưng lại không đủ lượng ăn vặt cung cấp cho nhu cầu của khách hàng.

Nhìn vào biểu đồ trên ta thấy biến Giai Khat và t có mối tương quan nghịch, biến An Vat và t có mối tương quan thuận. Tuy nhiên, mối tương quan giữa các biến trên yếu.



Hình 3: Biểu đồ tán xạ và hệ số tương quan của các biến Giai Khat, An Vat và t

Dùng phần mềm R, phương trình hồi quy tuyến tính mẫu giữa các biến Giai Khat, An Vat và biến t nhận được lần lượt như sau:

Phương trình 1: $Giai\ Khat = 37.44852 - 0.07221t,$

Phương trình 2: $An\ Vat = 22.36141 + 0.06761t.$

Cả hai mô hình trên đều có p-value < 0.05 nhưng có hệ số xác định không cao nên chỉ có ý nghĩa tương đối về mặt thống kê, chưa có độ tin cậy để sử dụng dự báo từ hai mô hình này. Tuy nhiên, chúng ta có thể sử dụng chúng để phân tích, quan sát xu hướng. Phương trình hồi quy cho thấy số lượng giải khát bán ra có xu hướng giảm, ngược lại số lượng ăn vặt bán ra lại có xu hướng tăng. Hai điều này dẫn đến một số câu hỏi đặt ra cho người quản lý: liệu có phải khách hàng hay dùng các mặt hàng giải khát đang bị giảm sút (do chất lượng sản phẩm không tốt), căng tin đang bị mất đi một số lượng khách hàng này? Và số lượng khách ăn vặt đang tăng lên, có phải là do mặt hàng này hấp dẫn hơn? Nhưng mặt khác ta thấy hai hệ số xu hướng của hai phương trình này có trị số gần như đối nhau, đặt ra khả năng là các khách hàng quen thuộc vẫn trung thành với căn tin, nhưng có một số chuyển từ chọn lựa sản phẩm giải khát sang chọn

lựa các món hàng ăn vặt. Điều này được kiểm chứng qua hai phương trình sau:

Phương trình 3: $AV\&GK = 59.80966 - 0.0046t,$

Phương trình 4: $Tongdoanhthu = 1709110 + 602.3t.$

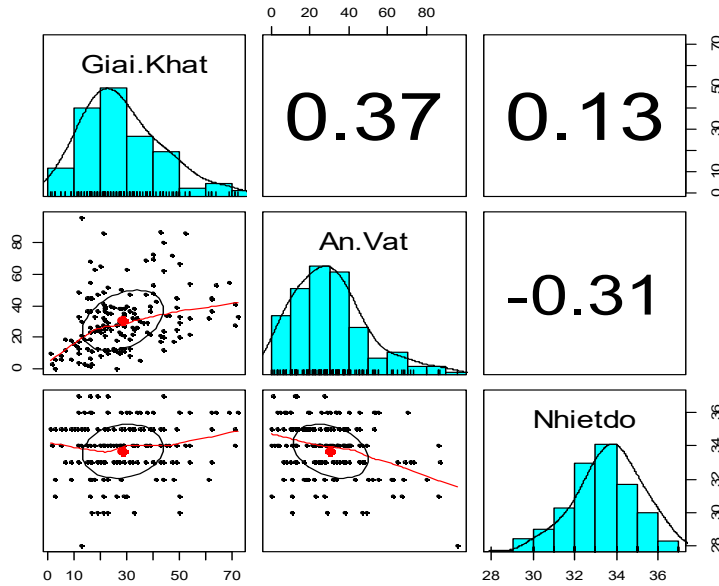
Khi hai biến An Vat và Giai Khat kết hợp với nhau thì xu hướng không có ý nghĩa về mặt thống kê (p-value = 0.8788 ở **Phương trình 3**) và nó cũng không có ý nghĩa về mặt buôn bán (p-value = 0.458 ở **Phương trình 4**). Từ các kết quả thu được ở đây giúp ta nhận ra rằng lượng giải khát bán ra giảm nhưng bù vào đó là mức tiêu thụ của món ăn vặt lại tăng và ảnh hưởng cuối cùng trong toàn bộ việc buôn bán theo thời gian là không đáng kể. Chúng ta thử tìm lý do của sự biến động trái chiều của hai biến thống kê này qua một nhân tố có thể có ảnh hưởng, đó là sự biến đổi thời tiết, khí hậu trong thời gian khảo sát, được xem xét cụ thể qua biến Nhiệt độ.

Biểu đồ phân tích sự tương quan giữa các biến An Vat, Giai Khat và Nhiệt độ nhận được ở Hình 4.

Qua biểu đồ trên, ta thấy biến Giai Khat và Nhiệt độ có mối tương quan thuận, hai biến An Vat và Nhiệt độ lại có tương quan nghịch. Từ đó ta có

thể khẳng định sự phù hợp giữa thực tế quan sát và phân tích thống kê xu hướng: khi nhiệt độ có xu hướng giảm (xem ở mục 3.2.1) thì số lượng giải khát bán ra cũng giảm, ngược lại số lượng bán được của mặt hàng ăn vặt thì tăng lên. Tức là có sự

chuyển đổi (do sự tác động của việc nhiệt độ giảm) từ sự chọn lựa sản phẩm giải khát sang các mặt hàng ăn vặt, còn về số lượng khách hàng trung thành của căn tin vẫn ổn định. Thông tin phân tích này rất có ý nghĩa đối với người quản lý.



Hình 4: Biểu đồ tán xạ và hệ số tương quan của các biến Giai Khat, An Vat và Nhiệt độ

3.3 Phân tích hồi quy đa biến với biến giả

Bước tiếp theo có thể có trong dự báo doanh số bán hàng là việc sử dụng hồi quy đa biến với việc tiên đoán bao gồm thời gian, các ngày trong tuần, và có thể là nhiệt độ. Trong phương pháp này, chúng tôi sử dụng sáu biến giả đại diện cho thứ hai, thứ ba, thứ tư, thứ năm, thứ sáu và thứ bảy. Chủ nhật được chọn là trường hợp cơ sở, vì vậy nó được đại diện bởi các số 0 cho tất cả biến giả (sáu ngày khác trong tuần). Sau đó, mỗi thứ hai được cho một giá trị "1" cho biến giả thứ hai và một giá trị "0" cho mỗi biến giả khác. Tương tự như vậy,

các thứ 3 được biểu thị bởi "1" cho các biến giả thứ 3 và "0" cho các biến giả khác...

Phương trình hồi quy đa biến để dự báo doanh thu dựa vào biến t và các biến Thu:

$$\text{Tongdoanhthu} = 1266816.4 + 616.8t + 532326.1\text{Thuhai} + 538167.1\text{Thuba} + 551022.7\text{Thutu} + 746345.7\text{Thunam} + 593288.9\text{Thusau} - 40979.9\text{Thubay}.$$

Giá trị p-value = 0.002173 < 0.05 nên mô hình có ý nghĩa về mặt thống kê.

Bảng 4: Tóm tắt kết quả dự báo bằng phân tích hồi quy với biến giả

t	Thứ	Doanh thu (đồng) dự báo từ mô hình hồi quy với biến giả	Doanh thu (đồng) dự báo từ mô hình ARIMA
243	Thứ năm	2163044.5	2985823
244	Thứ sáu	2010604.7	2762798
245	Thứ bảy	1376952.5	2684118
246	Chủ nhật	1418549.2	2656361
247	Thứ hai	1951492.1	2646568
248	Thứ ba	1957949.9	2643114
249	Thứ tư	1971422.3	2641895

Trong phương trình này, hệ số của biến thời gian t là 616.8, nó cho biết một xu hướng tăng trong tổng doanh thu. Các hệ số của các biến giả ngày trong tuần đại diện cho số tiền trung bình mà

mỗi ngày trong tuần khác với ngày cơ sở. Ví dụ, hệ số của biến thứ 2 là 532326.1, cho biết rằng trung bình doanh số bán hàng của thứ 2 là khoảng

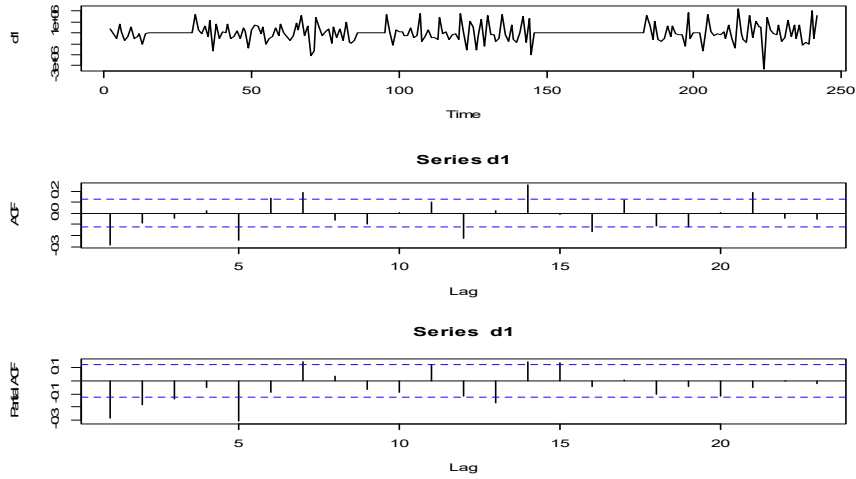
532326.1 đồng nhiều hơn doanh số bán hàng vào ngày chủ nhật.

Thực hiện báo doanh thu cho những ngày kế tiếp từ mô hình này, ví dụ khi $t = 243$ (Thứ năm) thì $Tongdoanhthu = 1266816.4 + 616.8(243) + 746345.7 = 2163044.5$, và tính tương tự cho

những ngày khác, ta thu được kết quả sau ở cột 3, Bảng 4.

3.4 Phân tích chuỗi thời gian

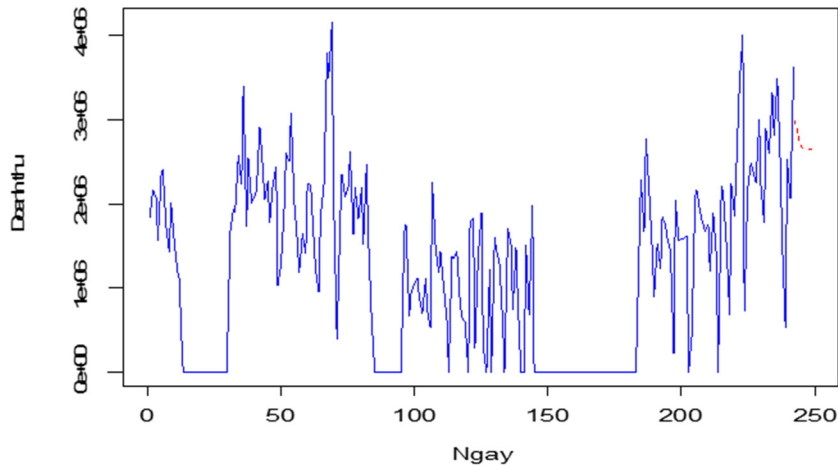
Do số liệu ban đầu của doanh thu chưa có tính dừng nên ta lấy sai phân bậc 1 và thu được kết quả như sau:



Hình 5: đồ thị plot, acf và pacf sau khi lấy sai phân bậc 1

Nhìn đồ thị ta thấy có tính dừng nên chọn $p = 3$, $d = 1$, $q = 1$. Tìm mô hình phù hợp để dự báo và kết quả chọn được là ARIMA(1,1,1) có chỉ số AIC

nhỏ nhất nên chúng tôi chọn mô hình này. Kết quả dự báo thu được ở cột 4, Bảng 4 và đồ thị dự báo được biểu diễn ở Hình 5.



Hình 6: Đồ thị dự báo cho doanh thu theo mô hình chuỗi thời gian

Nhìn vào đồ thị dự báo (cung đường vẽ đứt khúc), ta thấy doanh thu của quán trong những ngày dự báo tiếp theo có xu hướng giảm (sau đó lại quay về một mức ổn định nếu thực hiện dự báo tiếp tục). Điều này không hợp lý lắm với môi trường kinh doanh trong trường học với khách hàng chủ yếu là sinh viên có số lượng biến thiên thay đổi

phụ thuộc vào lịch học của các ngày khác nhau trong tuần.

4 KẾT LUẬN

Hai mô hình thống kê đã được sử dụng trong bài viết. Kết quả chứng tỏ mô hình ARIMA theo

chuỗi thời gian được áp dụng không hợp lý lắm trong trường hợp này. Mô hình hồi quy với biến giả thể hiện là mô hình phù hợp hơn. Qua đó ta có thể đưa ra kết luận rằng để sử dụng khảo sát cho dạng “dữ liệu căn tin” (“café data” – theo cách gọi của DePaolo *et al.*, 2011, [4]) trong môi trường hoạt động như trường học nói riêng hay cho những trường hợp khác mà dữ liệu phụ thuộc theo lịch học, lịch làm việc theo tuần hay theo mùa... thì ta nên ưu tiên xem xét sử dụng mô hình hồi quy với “biến giả mùa”. Đây là dạng mô hình có thể giải thích và dự báo hợp lý cho những biến thiên có tính chu kỳ của dạng dữ liệu đặc thù này.

Đối với bộ dữ liệu được khảo sát trong bài, các kết quả phân tích có thể giúp cho người quản lý căn tin vạch ra chiến lược kinh doanh tốt hơn qua việc bố trí số lượng nhân viên mỗi ngày một cách hợp lý cũng như việc tăng, giảm số lượng mặt hàng sao cho phù hợp để đạt được mức doanh thu và lợi nhuận cao nhất có thể. Qua phân tích hiện trạng cho thấy tình hình buôn bán của căn tin không ổn định, có sự chênh lệch giữa các học kỳ, đáng chú ý hơn là có sự khác biệt lớn về doanh thu của các ngày trong tuần qua đó ta thấy được sự phụ thuộc của doanh thu vào lịch học sinh viên là rất lớn.

Ở khía cạnh khác, ta thấy lượng bán của các mặt hàng ăn vặt có xu hướng tăng theo thời gian trong khi đó thì phía giải khát lại có xu hướng giảm. Nếu xem xét một cách riêng rẽ thì có thể cho là số lượng khách hàng của nhóm ăn vặt ngày càng tăng (do các món ăn ngon, hấp dẫn), còn lượng khách dùng giải khát đang giảm đi (do sản phẩm giải khát kém ngon). Tuy nhiên, điều thú vị ở đây là khi gộp hai mặt hàng này lại thì xu hướng lại không có ý nghĩa về mặt thống kê và yếu tố thời gian không làm ảnh hưởng đến toàn bộ việc buôn bán. Phân tích thống kê trong mô hình xu thế được áp dụng ở đây đã tìm ra được một giải thích hợp lý.

Đó là do sự tác động của nhân tố thời tiết khí hậu, cụ thể là biến nhiệt độ. Khảo sát cho thấy nhiệt độ có xu hướng giảm dần trong thời gian thu thập dữ liệu. Và do thời tiết mát mẻ như vậy nên khách hàng có xu hướng chuyển đổi lựa chọn từ các mặt hàng giải khát sang lựa chọn các món ăn vặt nhiều hơn, có nghĩa là lượng khách hàng trung thành của căn tin là không biến thiên. Rõ ràng sự áp dụng phân tích thống kê một cách hợp lý trong trường hợp này đã cung cấp cho chúng ta những xem xét, giải thích rất khoa học, rất thú vị và rất có ích trong quản lý.

LỜI CẢM ƠN

Các tác giả trân trọng cảm ơn căn tin Văn phòng Đoàn Trường Đại học Cần Thơ đã cho phép chúng tôi khai thác và sử dụng dữ liệu của căn tin. Nhờ có những dữ liệu này, chúng tôi mới có thể thực hiện được các áp dụng phù hợp với mục tiêu của bài báo và phù hợp với điều kiện nghiên cứu của mình.

TÀI LIỆU THAM KHẢO

- Aczet, A.D., Sounderpandian, J., 2009. Complete Business Statistics (7th ed). New York: McGraw-Hill/Irwin, 523 pages.
- Anderson, D.R., Sweeney, D.J., Williams, T.A., 2011. Statistics for Business and Economics (11th ed). Mason, OH: South-Western, 631 pages.
- Cleff, T., 2011. Exploratory Data Analysis in Business and Economics. Springer, 347 pages.
- Concetta A.D., David F.R., 2011. Café data, Journal of Statistics Education, 19(1): 123-135.
- Groebner, D.F., Shannon, P.W., Fry, P.C., Smith, K.D., 2011. Business Statistics: A Decision-Making Approach (8th ed). Upper Saddle River, NJ: Prentice-Hall, 467 pages.
- Trần Hùng Thao, 2013. Toán tài chính căn bản. NXB Văn hóa thông tin.