

## Chương Trình Giảng Dạy Kinh tế Fulbright

Học kỳ Thu năm 2013

### Các Phương Pháp Phân Tích Định Lượng

#### Gợi ý giải Bài tập 6

#### HỒI QUY ĐƠN BIẾN

**Ngày Phát: Thứ ba 5/11/2013**

**Ngày Nộp: Thứ ba 12/11/2013**

Bản in nộp lúc **8h20 sáng**, tại Hộp nộp bài tập trong phòng Lab

Bản điện tử gửi theo hướng dẫn của trợ giảng.

---

#### **Bài 1: (25 điểm)**

Trong thuật ngữ “hồi quy tuyến tính”, cụm từ “tuyến tính” được hiểu theo nghĩa nào?

Hồi quy tuyến tính có nghĩa là hồi quy tuyến tính theo các thông số (các  $\beta$ ), nghĩa là các thông số chỉ có số bậc bằng 1. Nó có thể tuyến tính hoặc không tuyến tính theo các biến giải thích.

Các mô hình liệt kê dưới đây có phải là mô hình tuyến tính không

Đối với các mô hình “không tuyến tính”, hãy thử dùng các phép biến đổi toán học thích hợp để chuyển chúng về dạng tuyến tính.

a.  $Y_i = \beta_1 + \beta_2 \frac{1}{\ln X_i} + u_i$

Đây là mô hình tuyến tính theo tham số, phi tuyến theo biến số.

b.  $Y_i = \beta_1 + \beta_2 \ln(X_i) + u_i$

Đây là mô hình tuyến tính theo tham số, phi tuyến theo biến số.

c.  $Y_i = \beta_1 + e^{\beta_2 / X_i} + u_i$

Đây không là mô hình tuyến tính theo tham số. Ta biến đổi về mô hình tuyến tính như sau:<sup>1</sup>

$$Y_i = \beta_1 + e^{\beta_2 / X_i} + u_i$$

$$\Leftrightarrow Y_i - \beta_1 - u_i = e^{\beta_2 / X_i}$$

$$\Leftrightarrow \ln(Y_i - \beta_1 - u_i) = \ln(e^{\beta_2 / X_i})$$

$$\Leftrightarrow \ln(Y_i - \beta_1 - u_i) = \beta_2 / X_i$$

$$\Leftrightarrow Y_i' = \beta_2 X_i' \text{ với } Y_i' = \ln(Y_i - \beta_1 - u_i) \text{ và } X_i' = 1 / X_i$$

---

<sup>1</sup> Vì  $u_i$  là sai số nên khi biến đổi có thể tạm loại bỏ ra khỏi mô hình và cộng thêm trở lại ( $v_i$ ) sau khi đã biến đổi xong. Học viên cũng có thể thêm vào các điều kiện để các phép biến đổi có ý nghĩa về mặt toán học.

d.  $Y_i = e^{\beta_1 + \beta_2 / X_i + u_i}$

Đây không là mô hình tuyến tính theo tham số. Ta biến đổi về mô hình tuyến tính như sau:

$$Y_i = e^{\beta_1 + \beta_2 / X_i + u_i}$$

$$\Leftrightarrow \ln(Y_i) = \ln(e^{\beta_1 + \beta_2 / X_i + u_i})$$

$$\Leftrightarrow \ln(Y_i) = \beta_1 + \beta_2 / X_i + u_i$$

$$\Leftrightarrow Y_i' = \beta_1 + \beta_2 X_i' + u_i \text{ với } Y_i' = \ln(Y_i) \text{ và } X_i' = 1 / X_i$$

e.  $Y_i = \frac{1}{1 + e^{\beta_1 + \beta_2 X_i}} + u_i$

Đây không là mô hình tuyến tính theo tham số. Ta biến đổi về mô hình tuyến tính như sau:

$$Y_i = \frac{1}{1 + e^{\beta_1 + \beta_2 X_i}} + u_i$$

$$\Leftrightarrow \frac{1}{Y_i - u_i} = 1 + e^{\beta_1 + \beta_2 X_i}$$

$$\Leftrightarrow \ln\left(\frac{1}{Y_i - u_i} - 1\right) = \ln(e^{\beta_1 + \beta_2 X_i})$$

$$\Leftrightarrow \ln\left(\frac{1}{Y_i - u_i} - 1\right) = \beta_1 + \beta_2 X_i$$

Đặt  $Y_i' = \ln\left(\frac{1}{Y_i - u_i} - 1\right) \Rightarrow$  ta chuyển được mô hình ban đầu về mô hình tuyến tính theo tham số dưới đây:  $Y_i' = \beta_1 + \beta_2 X_i$

f.  $\ln Y_i = \frac{X_i}{\beta_1 + \beta_2 X_i} + u_i$

Đây không là mô hình tuyến tính theo tham số. Ta biến đổi về mô hình tuyến tính như sau:

$$\ln Y_i = \frac{X_i}{\beta_1 + \beta_2 X_i} + u_i$$

$$\Leftrightarrow \ln Y_i - u_i = \frac{X_i}{\beta_1 + \beta_2 X_i}$$

$$\Leftrightarrow \frac{1}{\ln Y_i - u_i} = \frac{\beta_1 + \beta_2 X_i}{X_i}$$

$$\Leftrightarrow \frac{1}{\ln Y_i - u_i} = \beta_2 + \frac{\beta_1}{X_i}$$

$$\Leftrightarrow Y_i' = \beta_2 + \beta_1 X_i' \text{ với } Y_i' = \frac{1}{\ln Y_i - u_i} \text{ và } X_i' = \frac{1}{X_i}$$

g.  $Y_i = \beta_1 + \ln(\beta_2)X_i + u_i$

Đây chưa là mô hình tuyến tính theo tham số.

Ta đặt  $\beta_2' = \ln(\beta_2) \Rightarrow$  Ta có phương trình mới tuyến tính theo tham số:  $Y_i = \beta_1 + \beta_2'X_i + u_i$

h.  $\ln Y_i = \beta_1 + \beta_2 e^{\beta_2 X_i} + u_i$

Đây chưa là mô hình tuyến tính theo tham số. Để chuyển về mô hình tuyến tính theo tham số ta thực hiện như sau:

Cách 1:

$$\ln Y_i = \beta_1 + \beta_2 e^{\beta_2 X_i} + u_i$$

$$\Leftrightarrow \ln Y_i - \beta_1 - u_i = \beta_2 e^{\beta_2 X_i}$$

$$\Leftrightarrow \frac{\ln Y_i - \beta_1 - u_i}{\beta_2} = e^{\beta_2 X_i}$$

$$\Leftrightarrow \ln\left(\frac{\ln Y_i - \beta_1 - u_i}{\beta_2}\right) = \beta_2 X_i$$

$$\Leftrightarrow \left(\ln\left(\frac{\ln Y_i - \beta_1 - u_i}{\beta_2}\right)\right)^{1/X_i} = \beta_2$$

$$\Leftrightarrow Y_i' = \beta_2 \text{ với } Y_i' = \left(\ln\left(\frac{\ln Y_i - \beta_1 - u_i}{\beta_2}\right)\right)^{1/X_i}$$

$$\Leftrightarrow Y_i' = \beta_2$$

Cách 2:

$$\ln Y_i = \beta_1 + \beta_2 e^{\beta_2 X_i} + u_i$$

$$\Leftrightarrow \frac{\ln Y_i - \beta_1 - u_i}{\beta_2} = e^{\beta_2 X_i}$$

$$\text{Lấy ln của 2 vế ta được } \ln\left(\frac{\ln Y_i - \beta_1 - u_i}{\beta_2}\right) = \ln(e^{\beta_2 X_i}) = \beta_2 X_i$$

$$\text{Tiếp tục lấy ln 2 vế ta được } \ln\left(\ln\left(\frac{\ln Y_i - \beta_1 - u_i}{\beta_2}\right)\right) = \ln(\beta_2 X_i) = X_i \ln(\beta_2)$$

$$\text{Đặt } Y_i' = \ln\left(\ln\left(\frac{\ln Y_i - \beta_1 - u_i}{\beta_2}\right)\right) \text{ và } \alpha = \ln(\beta_2)$$

$$\Rightarrow \text{Ta có mô hình tuyến tính } Y_i' = \alpha X_i$$

i.  $\ln Y_i = \beta_1 + \beta_2 / X_i + u_i$

Đây là mô hình tuyến tính theo tham số nhưng chưa tuyến tính theo biến số. Để chuyển về mô hình tuyến tính theo cả tham số và biến số ta thực hiện như sau:

$$\ln Y_i = \beta_1 + \beta_2 / X_i + u_i$$

$$\Leftrightarrow Y_i' = \beta_1 + \beta_2 X_i' + u_i \text{ với } Y_i' = \ln Y_i \text{ và } X_i' = 1/X_i$$

j.  $\ln Y_i = \beta_1 + \ln(\beta_2 X_i) + u_i$

Đây chưa là mô hình tuyến tính theo tham số. Để chuyển về mô hình tuyến tính theo tham số ta thực hiện như sau:

$$\ln Y_i = \beta_1 + \ln(\beta_2 X_i) + u_i$$

$$\Leftrightarrow \ln Y_i = \beta_1 + \ln(\beta_2) + \ln X_i + u_i$$

$$\Leftrightarrow \ln Y_i = \beta_1 + \beta_2' + \ln X_i + u_i \text{ với } \beta_2' = \ln(\beta_2)$$

## Bài 2: (25 điểm)

Biến KIDS biểu thị số con sinh ra của một người phụ nữ, biến EDUC biểu thị số năm đi học của người phụ nữ này. Mô hình đơn giản về mối quan hệ giữa số con và số năm đi học như sau:

$$KIDS_i = \beta_1 + \beta_2 EDUC_i + u_i$$

Trong đó  $u_i$  là sai số ngẫu nhiên.

- a.  $u_i$  chứa những loại yếu tố nào? Những yếu tố này có khuynh hướng tương quan với số năm đi học hay không?

| Các yếu tố có thể có của $u_i$ | Tương quan với EDUC |
|--------------------------------|---------------------|
| Sức khỏe                       | Có                  |
| Thu nhập                       | Có                  |
| Nơi sống                       | Có                  |
| Công việc                      | Có                  |
| Luật pháp                      | Có                  |
| Sở thích                       | Có                  |
| Tôn giáo                       | Có                  |
| ...                            | ...                 |

Ngoài ra  $u_i$  còn có thể hình thành do sai số đo lường, dạng hàm sai, dữ liệu không có sẵn...

- b. Liệu phân tích hồi quy đơn nói trên có cho thấy tác động (các yếu tố khác không đổi) của số năm đi học đến số con sinh ra hay không? Giải thích lý do.

Phân tích hồi quy đơn nói trên **khó hay thậm chí là không thể** cho thấy tác động của số năm đi học đến số con sinh ra bởi trong mô hình, phần dư  $u_i$  còn chứa rất nhiều yếu tố khác nữa như đã liệt kê ở trên có tương quan với EDUC, tức là biến độc lập và phần dư còn có tương quan với nhau. Khi EDUC thay đổi thì phần dư cũng bị tác động và thay đổi theo một cách không ngẫu nhiên, do đó không thể giải thích được sự thay đổi của KIDS.

## Bài 3: (25 điểm)

Hàm tiêu dùng tuyến tính thể hiện như sau:

$$CONS = \beta_1 + \beta_2 INC_i$$

Trong đó  $CONS_i$ : Tiêu dùng của hộ gia đình  $i$

$INC_i$ : Thu nhập của hộ gia đình  $i$

Khuynh hướng tiêu dùng trung bình APC sẽ là

$$CONS / INC = \hat{\beta}_1 / INC + \hat{\beta}_2.$$

Với bộ số liệu của 100 hộ gia đình về thu nhập và chi tiêu tiêu dùng hàng năm (tính bằng USD), ta có phương trình sau

$$CONS = -128,84 + 0,853INC$$

a. Giải thích ý nghĩa kinh tế của hệ số  $\beta_2$ , trong mô hình này khuynh hướng tiêu dùng cận biên MPC là bao nhiêu? Giải thích ý nghĩa kinh tế lượng của hệ số này.

- Ý nghĩa kinh tế: hệ số  $\beta_2$  chính là khuynh hướng tiêu dùng cận biên (MPC). Nó cho biết tác động biên của *Thu nhập của hộ gia đình (INC)* lên *Tiêu dùng của hộ gia đình (CONS)*, khi *Thu nhập của hộ gia đình* tăng thêm 1 đơn vị thì *Tiêu dùng của hộ gia đình* sẽ tăng  $\beta_2$  đơn vị.
- Trong mô hình này, Khuynh hướng tiêu dùng cận biên  $MPC = \beta_2 = 0,853$
- Giải thích ý nghĩa kinh tế lượng của hệ số này: Trong điều kiện các yếu tố khác không đổi, nếu Thu nhập của hộ gia đình tăng lên 1 USD thì trung bình tiêu dùng của hộ gia đình tăng lên 0,853 USD và ngược lại.

b. Chi tiêu tiêu dùng dự đoán của hộ gia đình có thu nhập \$30.000 là bao nhiêu?

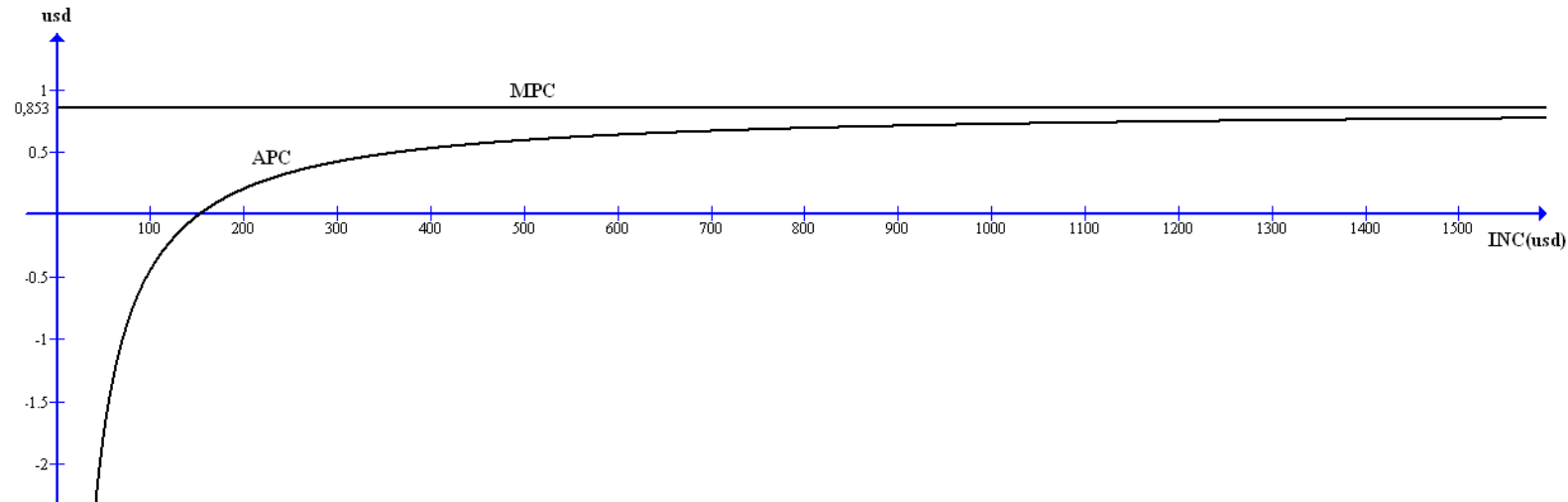
$$CONS = -128,84 + 0,853INC = -128,84 + 0,853 \cdot 30.000 = 25461,16 \text{ (USD)}$$

c. Với biến  $INC$  trên trục hoành hãy vẽ đường biểu diễn MPC và APC ước tính theo phương trình này.

Đường APC:  $CONS / INC = -128,84 / INC + 0,853$

Đường MPC:  $MPC = 0,853$

## ĐỒ THỊ ĐƯỜNG MPC VÀ ĐƯỜNG APC



### Bài 4: (25 điểm)

Tập đoàn du lịch Woody có một hệ thống gồm rất nhiều nhà hàng ở rất nhiều địa phương. Tập đoàn này đã thu thập dữ liệu mẫu gồm 33 nhà hàng. Thông tin của mỗi nhà hàng được lưu trữ trong file WOODY's DATA bao gồm các biến sau:

- Y: Số lượt khách hàng được phục vụ trong 1 năm (lượt khách)
  - I: Thu nhập trung bình một hộ của những hộ ở quanh nhà hàng trong phạm vi bán kính 3 dặm<sup>2</sup> (USD)
  - P: Số người dân sống xung quanh nhà hàng trong phạm vi bán kính 3 dặm (người)
  - N: Số đối thủ cạnh tranh trực tiếp với nhà hàng trong phạm vi bán kính 2 dặm
- a. Khung lý thuyết nào sẽ giúp bạn thiết lập mô hình về các yếu tố ảnh hưởng đến Y. Bạn kỳ vọng như thế nào về mối quan hệ giữa I và Y, giữa P và Y, giữa N và Y.
  - b. Hãy vẽ biểu đồ phân tán, tính hệ số tương quan của Y và I, Y và P, Y và N và cho biết nhận định của bạn về mối quan hệ giữa các biến này
  - c. Hãy sử dụng các công thức để tính toán các hệ số hồi quy trong hàm hồi quy tuyến tính  $Y=f(P)$

<sup>2</sup> Một dặm bằng 1.6 km

- d. Hãy sử dụng các công thức để tính toán phương sai, sai số chuẩn của các hệ số hồi quy trong hàm hồi quy tuyến tính  $Y=f(P)$
- e. Bạn hãy vẽ biểu đồ phân tán giữa phần dư của mô hình ở câu d (trục tung), và biến P (trục hoành)
- f. Hãy sử dụng công cụ Tools\Data Analysis\Regression (hoặc Stata) để tính toán kết quả các mô hình hồi quy tuyến tính  $Y=f(P)$ ,  $Y=f(N)$ ,  $Y=f(I)$ ; viết các phương trình và cho biết nhận xét của bạn.

**a)** Khung lý thuyết nào sẽ giúp bạn thiết lập mô hình về các yếu tố ảnh hưởng đến Y. Bạn kỳ vọng thế nào về mối quan hệ giữa I và Y, giữa P và Y, giữa N và Y.

Dựa vào khung lý thuyết cung cầu để thiết lập mô hình về các yếu tố ảnh hưởng đến Y. Lý thuyết cung cầu cho thấy cầu của một hàng hóa phụ thuộc vào thu nhập của người tiêu dùng, thị hiếu của người tiêu dùng, giá kỳ vọng, số người mua, giá hàng thay thế, giá hàng hóa bổ sung.

\*Kỳ vọng về mối quan hệ giữa I và Y:

I là thu nhập trung bình một hộ cả những hộ ở quanh nhà hàng trong phạm vi bán kính 3 dặm, khi thu nhập của người dân tăng lên, người ta có xu hướng đi ăn ở nhà hàng nhiều hơn. Do vậy, kỳ vọng về mối quan hệ giữa I và Y là đồng biến.

\*Kỳ vọng về mối quan hệ giữa P và Y:

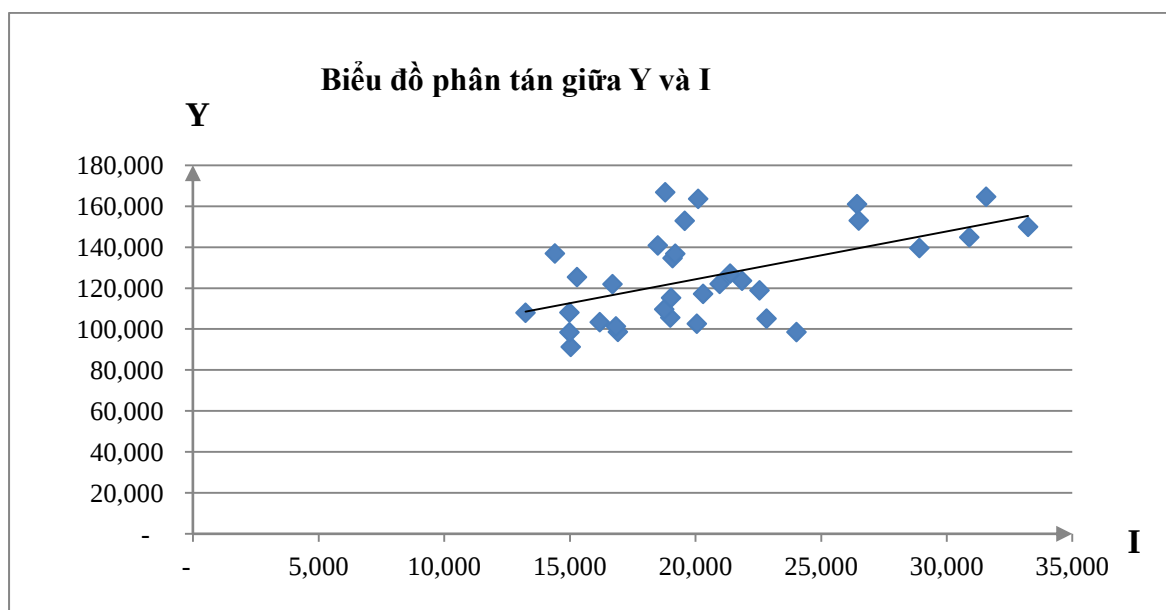
P là số người dân sống xung quanh nhà hàng trong phạm vi bán kính 3 dặm, khi số người dân tăng lên, có xu hướng số lượng khách đi ăn ở nhà hàng nhiều hơn. Do vậy, kỳ vọng về mối quan hệ giữa P và Y là đồng biến.

\*Kỳ vọng về mối quan hệ giữa N và Y:

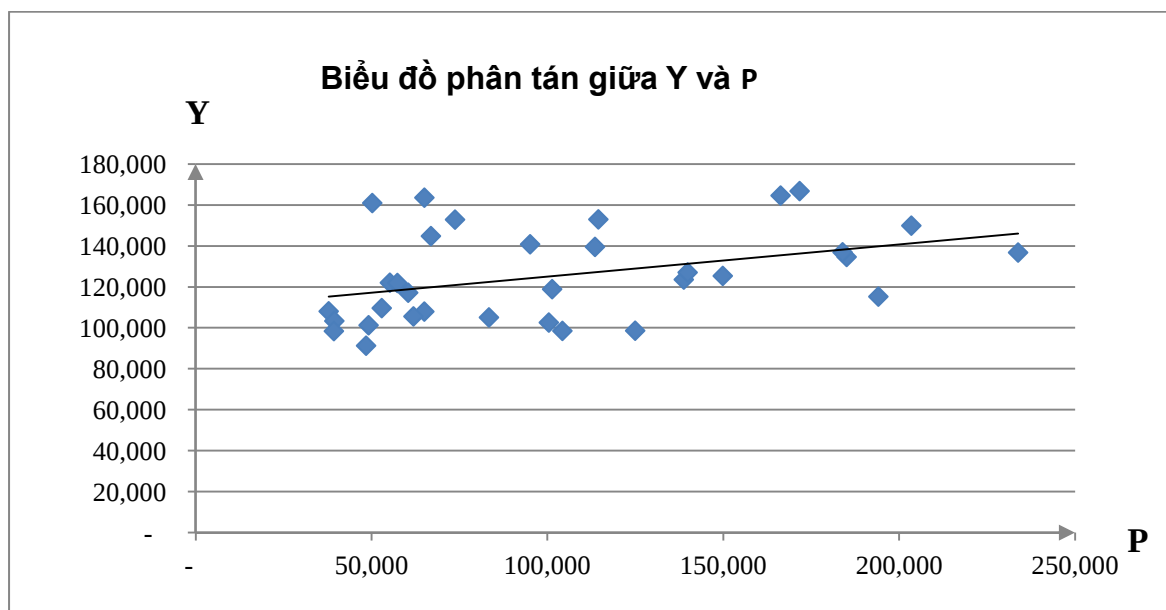
N là số đối thủ cạnh tranh trực tiếp với nhà hàng trong phạm vi bán kính 2 dặm, khi số đối thủ tăng lên, có xu hướng số lượng khách đi ăn ở nhà hàng này sẽ ít đi. Do vậy, kỳ vọng về mối quan hệ giữa N và Y là nghịch biến.

**b)** Hãy vẽ biểu đồ phân tán, tính hệ số tương quan của Y và I, Y và P, Y và N và cho biết nhận định của bạn về mối quan hệ giữa các biến này

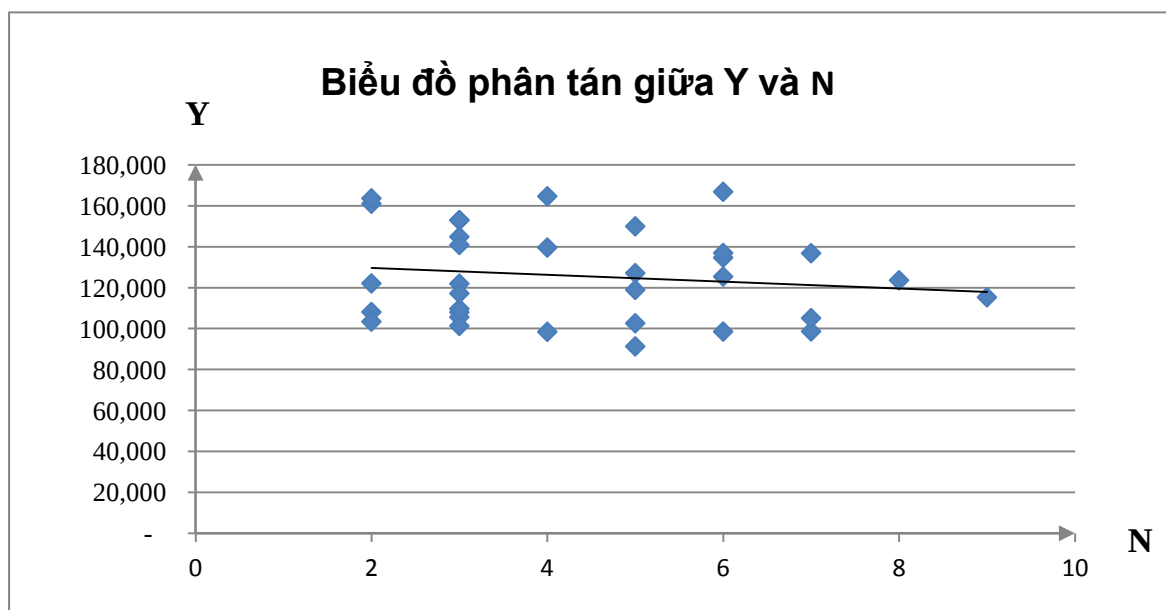
\*Biểu đồ phân tán của Y và I



\*Biểu đồ phân tán Y và P



\*Biểu đồ phân tán của Y và N



\*Tính hệ số tương quan:

Trong Excel 2007, ta dùng lệnh Correlation (trong mục Data -> Data Annalysis ->Correlation như sau:

|    | A       | B       | C       | D | E      |    |
|----|---------|---------|---------|---|--------|----|
| 2  | Nhahang | Y       | P       | N | I      | DO |
| 3  | 1       | 107,919 | 65,044  | 3 | 13,240 |    |
| 4  | 2       | 118,866 | 101,376 | 5 | 22,554 |    |
| 5  | 3       | 98,579  | 124,989 | 7 | 16,916 |    |
| 6  | 4       | 122,015 | 55,249  | 2 | 20,967 |    |
| 7  | 5       | 152,827 | 73,775  | 3 | 19,576 |    |
| 8  | 6       | 91,259  | 48,484  | 5 | 15,039 |    |
| 9  | 7       | 123,550 | 138,809 | 8 | 21,857 |    |
| 10 | 8       | 160,931 | 50,244  | 2 | 26,435 |    |
| 11 | 9       |         |         |   | 24,024 |    |
| 12 | 10      |         |         |   | 14,987 |    |
| 13 | 11      |         |         |   | 30,902 |    |
| 14 | 12      |         |         |   | 31,573 |    |
| 15 | 13      |         |         |   | 19,001 |    |
| 16 | 14      |         |         |   | 20,058 |    |
| 17 | 15      |         |         |   | 16,194 |    |
| 18 | 16      |         |         |   | 21,384 |    |
| 19 | 17      |         |         |   | 18,800 |    |
| 20 | 18      |         |         |   | 15,289 |    |
| 21 | 19      |         |         |   | 16,702 |    |
| 22 | 20      | 134,594 | 185,105 | 6 | 19,093 |    |
| 23 | 21      | 152,937 | 114,520 | 3 | 26,502 |    |

Ta được kết quả như sau:

|   | Y           | P           | N            | I |
|---|-------------|-------------|--------------|---|
| Y | 1           |             |              |   |
| P | 0,392567733 | 1           |              |   |
| N | -0,14422464 | 0,726250705 | 1            |   |
| I | 0,537022014 | 0,245197642 | -0,031534052 | 1 |

Vậy, hệ số tương quan giữa Y và I là 0,537022014

hệ số tương quan giữa Y và P là 0,392567733

hệ số tương quan giữa Y và N là -0,14422464

Nhận định của bạn về mối quan hệ giữa các biến này:<sup>3</sup>

Nhận định về mối quan hệ Y và I: Hệ số tương quan giữa hai biến là 0,537022014 cho biết số lượt khách hàng được phục vụ trong 1 năm (Y) và thu nhập trung bình một hộ của những hộ ở quanh nhà hàng trong phạm vi bán kính 3 dặm (I) có mối quan hệ đồng biến và tương đối chặt với nhau.

Nhận định về mối quan hệ Y và P: Hệ số tương quan giữa hai biến là 0,392567733 cho biết số lượt khách hàng được phục vụ trong 1 năm (Y) và số người dân sống xung quanh nhà hàng trong phạm vi bán kính 3 dặm (P) có mối quan hệ đồng biến và tương đối không chặt với nhau

Nhận định về mối quan hệ Y và N: Hệ số tương quan giữa hai biến là -0,14422464 cho biết giữa số lượt khách hàng được phục vụ trong 1 năm (Y) và Số đối thủ cạnh tranh trực tiếp với nhà hàng trong phạm vi bán kính 2 dặm (N) có quan hệ nghịch biến và tương quan yếu.

c) Hãy sử dụng công thức để tính toán các hệ số hồi quy trong hàm hồi quy tuyến tính  $Y = f(P)$

<sup>3</sup> Thông thường, nếu tương quan giữa hai biến lớn hơn 0,8 được gọi là tương quan mạnh; từ 0,4->0,8 là tương quan trung bình và nhỏ hơn 0,4 là tương quan yếu.

Áp dụng công thức như sau:

$$\hat{\beta}_2 = \frac{\sum (P_i - \bar{P})(Y_i - \bar{Y})}{\sum (P_i - \bar{P})^2}$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{P}$$

Trong đó  $\bar{P} = 103.887$ ;  $\bar{Y} = 125,635$ ;

$$\Rightarrow \hat{\beta}_2 = 0,157380334$$

$$\Rightarrow \hat{\beta}_1 = 109.285$$

Vậy ta có hàm hồi quy của mẫu như sau:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i = 109.285 + 0,157380334 P_i$$

Có thể ước lượng hàm hồi quy tuyến tính tổng thể  $Y=f(P)$  như sau:

$$Y_i = 109.285 + 0,157380334 X_i + \hat{u}_i$$

d) Tính toán phương sai, sai số chuẩn của các hệ số hồi quy trong hàm hồi quy tuyến tính  $Y = f(P)$

|    | A       | B       | C                      | D                     | E                     | F                     | G                     | H                | I           | J                     |
|----|---------|---------|------------------------|-----------------------|-----------------------|-----------------------|-----------------------|------------------|-------------|-----------------------|
| 1  |         |         |                        |                       |                       |                       |                       |                  |             |                       |
| 2  | Y       | P       | $P^2$                  | $P_i = P_i - \bar{P}$ | $Y_i = Y_i - \bar{Y}$ | $P_i Y_i$             | $P_i^2$               | $Y_i^2$          | $\hat{u}_i$ | $\hat{u}_i^2$         |
| 21 | 121,886 | 57,386  | 3,293,152,996          | (46,501)              | (3,749)               | 174,315,748           | 2,162,388,093         | 118,316          | 3,570       | 12,743,566            |
| 22 | 134,594 | 185,105 | 34,263,861,025         | 81,218                | 8,959                 | 727,659,713           | 6,596,284,767         | 138,417          | (3,823)     | 14,612,620            |
| 23 | 152,937 | 114,520 | 13,114,830,400         | 10,633                | 27,302                | 290,293,117           | 113,050,378           | 127,308          | 25,629      | 656,847,955           |
| 24 | 109,622 | 52,933  | 2,801,902,489          | (50,954)              | (16,013)              | 815,914,093           | 2,596,359,526         | 117,615          | (7,993)     | 63,893,999            |
| 25 | 149,884 | 203,500 | 41,412,250,000         | 99,613                | 24,249                | 2,415,543,121         | 9,922,653,175         | 141,312          | 8,572       | 73,485,065            |
| 26 | 98,388  | 39,334  | 1,547,163,556          | (64,553)              | (27,247)              | 1,758,863,372         | 4,167,152,406         | 115,475          | (17,087)    | 291,970,936           |
| 27 | 140,791 | 95,120  | 9,047,814,400          | (8,767)               | 15,156                | (132,883,454)         | 76,868,791            | 124,255          | 16,536      | 273,446,692           |
| 28 | 101,260 | 49,200  | 2,420,640,000          | (54,687)              | (24,375)              | 1,332,985,900         | 2,990,720,999         | 117,028          | (15,768)    | 248,625,769           |
| 29 | 139,517 | 113,566 | 12,897,236,356         | 9,679                 | 13,882                | 134,360,960           | 93,673,656            | 127,158          | 12,359      | 152,749,478           |
| 30 | 115,236 | 194,125 | 37,684,515,625         | 90,238                | (10,399)              | (938,344,372)         | 8,142,809,141         | 139,836          | (24,600)    | 605,170,644           |
| 31 | 136,749 | 233,844 | 54,683,016,336         | 129,957               | 11,114                | 1,444,387,904         | 16,888,695,830        | 146,087          | (9,338)     | 87,202,088            |
| 32 | 105,067 | 83,416  | 6,958,229,056          | (20,471)              | (20,568)              | 421,049,436           | 419,081,692           | 122,413          | (17,346)    | 300,876,671           |
| 33 | 136,872 | 183,953 | 33,838,706,209         | 80,066                | 11,237                | 899,727,735           | 6,410,486,716         | 138,235          | (1,363)     | 1,858,706             |
| 34 | 117,146 | 60,457  | 3,655,048,849          | (43,430)              | (8,489)               | 368,664,277           | 1,886,207,014         | 118,800          | (1,654)     | 2,734,068             |
| 35 | 163,538 | 65,065  | 4,233,454,225          | (38,822)              | 37,903                | (1,471,503,937)       | 1,507,185,330         | 119,525          | 44,013      | 1,937,169,659         |
| 36 |         |         | <b>456,094,623,385</b> |                       |                       | <b>15,728,356,029</b> | <b>99,938,509,616</b> | <b>4,145,942</b> | -           | <b>13,586,849,956</b> |

$$\bullet \hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2} = \frac{13,586,849,956}{33-2} = 438,285,482 \text{ (lượt người)}^2$$

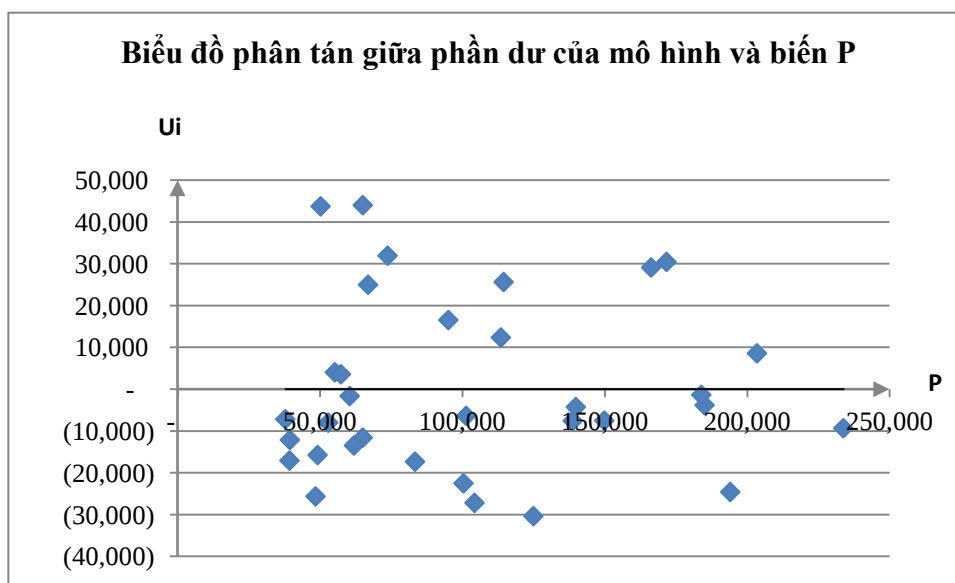
$$\text{Var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum P_i^2} = \frac{\hat{\sigma}^2}{\sum P_i^2} = \frac{438,285,482}{99,938,509,616} = 0,004385552 \text{ (lượt người)}^2$$

$$\Rightarrow \text{Se}(\hat{\beta}_2) = 0,0662235 \text{ (lượt người)}$$

$$\text{Var}(\hat{\beta}_1) = \frac{\sum P_i^2}{n \sum P_i^2} \sigma^2 = \frac{\sum P_i^2}{n \sum P_i^2} \hat{\sigma}^2 = \frac{456,094,623,385}{33 \times 99,938,509,616} \cdot 438,285,482 = 60,612,923.26 \text{ (lượt người)}^2$$

$$\Rightarrow \text{Se}(\hat{\beta}_1) = 7785,430191 \text{ (lượt người)}$$

e) Vẽ biểu đồ phân tán giữa phần dư của mô hình ở câu (d) và biến P



f) Hãy sử dụng công cụ *Tools\Data Analysis\Regression* (hoặc *Stata*) để tính toán kết quả các mô hình hồi quy tuyến tính  $Y=f(P)$ ,  $Y=f(N)$ ,  $Y=f(I)$ ; viết các phương trình và cho biết nhận xét của bạn.

Kết quả mô hình hồi quy tuyến tính  $Y=f(P)$ :

SUMMARY  
OUTPUT

| Regression Statistics |         |
|-----------------------|---------|
|                       | 0.39256 |
| Multiple R            | 7733    |
|                       | 0.15410 |
| R Square              | 9425    |
| Adjusted R Square     | 0.12682 |
|                       | 2633    |
|                       | 20935.2 |
| Standard Error        | 6886    |
| Observations          | 33      |

| ANOVA      |           |           |           |          |                       |
|------------|-----------|-----------|-----------|----------|-----------------------|
|            | <i>df</i> | <i>SS</i> | <i>MS</i> | <i>F</i> | <i>Significance F</i> |
| Regression | 1         | 24753339  | 247533    | 5.64776  | 0.023836              |
| Residual   | 31        | 13586849  | 438285    |          |                       |
| Total      | 32        | 16062183  |           |          |                       |

|  | <i>Coefficients</i> | <i>Standard Error</i> | <i>t Stat</i> | <i>P-value</i> | <i>Lower 95%</i> | <i>Upper 95%</i> | <i>Lower 95.0%</i> | <i>Upper 95.0%</i> |
|--|---------------------|-----------------------|---------------|----------------|------------------|------------------|--------------------|--------------------|
|--|---------------------|-----------------------|---------------|----------------|------------------|------------------|--------------------|--------------------|

|              |         |          |         |         |          |         |         |         |
|--------------|---------|----------|---------|---------|----------|---------|---------|---------|
|              | 109284. | 7785.430 | 14.0370 | 5.61538 | 93406.26 | 125163. | 93406.2 | 125163. |
| Intercept    | 759     | 191      | 8675    | E-15    | 949      | 2485    | 7       | 2       |
|              | 0.15738 | 0.066223 | 2.37650 | 0.02383 | 0.022316 | 0.29244 | 0.02231 | 0.29244 |
| X Variable 1 | 0334    | 497      | 2931    | 6465    | 623      | 4046    | 7       | 4       |

Vậy mô hình hồi quy tuyến tính  $Y=f(P)$  là  $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2.P_i = 109284,759 + 0,157380334 P_i$

Nhận xét: Ta thấy các ước lượng của hệ số beta đều có ý nghĩa thống kê ở mức ý nghĩa 5%. Các biến Y và P có mối quan hệ đồng biến, phù hợp với nhận định ban đầu về mối liên hệ giữa Y và P.

Với hệ số  $\beta_2 = 0,1574$  cho ta biết trong các điều kiện khác không đổi nếu số người dân sống xung quanh nhà hàng trong phạm vi bán kính 3 dặm tăng 1 người thì số lượt người được nhà hàng phục vụ trong một năm tăng 0,1574 lượt về mặt trung bình.<sup>4</sup>

#### Kết quả mô hình hồi quy tuyến tính $Y=f(N)$

SUMMARY  
OUTPUT

| <i>Regression Statistics</i> |         |
|------------------------------|---------|
|                              | 0.14422 |
| Multiple R                   | 464     |
|                              | 0.02080 |
| R Square                     | 0747    |
|                              | -       |
| Adjusted R Square            | 0.01078 |
|                              | 6326    |
|                              | 22524.5 |
| Standard Error               | 9477    |
| Observations                 | 33      |

#### ANOVA

|            | <i>df</i> | <i>SS</i> | <i>MS</i> | <i>F</i> | <i>Significance F</i> |
|------------|-----------|-----------|-----------|----------|-----------------------|
| Regression | 1         | 3341054   | 3341054   | 0.65852  | 0.42326               |
|            |           | 20.6      | 20.6      | 0878     | 9919                  |
| Residual   | 31        | 1572807   | 5073573   |          |                       |
|            |           | 8461      | 69.7      |          |                       |
| Total      | 32        | 1606218   |           |          |                       |
|            |           | 3882      |           |          |                       |

|              | <i>Coefficients</i> | <i>Standard Error</i> | <i>t Stat</i> | <i>P-value</i> | <i>Lower 95%</i> | <i>Upper 95%</i> | <i>Lower 95.0%</i> | <i>Upper 95.0%</i> |
|--------------|---------------------|-----------------------|---------------|----------------|------------------|------------------|--------------------|--------------------|
| Intercept    | 133031.9871         | 9923.289              | 13.4060       | 1.92166        | 112793.3048      | 153270.6694      | 112793.3           | 153270.7           |
|              | -                   | 504                   | 3709          | E-14           | -                | -                | -                  | -                  |
|              | 1683.54             | 2074.622              | 0.81149       | 0.42326        | 5914.76          | 2547.67          | -                  | 2547.67            |
| X Variable 1 | 1902                | 845                   | 2993          | 9919           | 3074             | 927              | 5914.76            | 9                  |

Vậy mô hình hồi quy tuyến tính  $Y=f(N)$  là  $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2.N_i = 133031.9871 - 1683.541902 N_i$

<sup>4</sup> Học viên có thể nhận xét thêm về hệ số R Square; Adjusted R Square;... cho từng mô hình.

Nhận xét: Ta thấy ước lượng của hệ số beta không có ý nghĩa thống kê ở mức ý nghĩa 5% do  $P_{\text{value}}$  có giá trị lớn, tuy nhiên trên thực tế và các cơ sở lý thuyết đều thấy số đối thủ cạnh tranh trực tiếp có ảnh hưởng đến số lượt khách hàng được phục vụ trong năm của nhà hàng, mối quan hệ này là nghịch biến và phù hợp với nhận định ban đầu về mối liên hệ giữa Y và N.

Với hệ số  $\beta_2 = -1.683,53$  cho ta biết trong các điều kiện khác không đổi nếu số đối thủ cạnh tranh trực tiếp với nhà hàng tăng lên 1 đơn vị thì số lượt người được nhà hàng phục vụ trong một năm sẽ giảm đi khoảng 1.684 lượt về mặt trung bình.

### Kết quả mô hình hồi quy tuyến tính $Y=f(I)$

#### SUMMARY OUTPUT

| Regression Statistics |         |
|-----------------------|---------|
|                       | 0.53702 |
| Multiple R            | 2014    |
|                       | 0.28839 |
| R Square              | 2643    |
| Adjusted R Square     | 0.26543 |
|                       | 7567    |
|                       | 19201.7 |
| Standard Error        | 8773    |
| Observations          | 33      |

#### ANOVA

|            | <i>df</i> | <i>SS</i> | <i>MS</i> | <i>F</i> | <i>Significance F</i> |
|------------|-----------|-----------|-----------|----------|-----------------------|
| Regression | 1         | 46322156  | 463221    | 12.5633  | 0.001271              |
|            |           | 66        | 5666      | 4951     | 677                   |
| Residual   | 31        | 11429968  | 368708    |          |                       |
|            |           | 216       | 652.1     |          |                       |
| Total      | 32        | 16062183  |           |          |                       |
|            |           | 882       |           |          |                       |

|              | <i>Coefficients</i> | <i>Standard Error</i> | <i>t Stat</i> | <i>P-value</i> | <i>Lower 95%</i> | <i>Upper 95%</i> | <i>Lower 95.0%</i> | <i>Upper 95.0%</i> |
|--------------|---------------------|-----------------------|---------------|----------------|------------------|------------------|--------------------|--------------------|
| Intercept    | 77543.4             | 13973.56              | 5.54929       | 4.44574        | 49044.19         | 106042.          |                    | 106042.            |
|              | 7617                | 728                   | 7084          | E-06           | 791              | 7544             | 49044.2            | 8                  |
|              | 2.33990             | 0.660155              | 3.54448       | 0.00127        | 0.993512         | 3.68630          | 0.99351            | 3.68630            |
| X Variable 1 | 7682                | 129                   | 1558          | 1677           | 424              | 2939             | 2                  | 3                  |

Vậy mô hình hồi quy tuyến tính  $Y=f(I)$  là  $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 \cdot I_i = 77543.47617 + 2.339907682 I_i$

Nhận xét Ta thấy các ước lượng của hệ số beta đều có ý nghĩa thống kê ở mức ý nghĩa 5%. Các biến Y và I có mối quan hệ đồng biến, phù hợp với nhận định ban đầu về mối liên hệ giữa Y và I.

Với hệ số  $\beta_2 = 2,3399$  cho ta biết trong các điều kiện khác không đổi thu nhập trung bình một hộ của những hộ ở quanh nhà hàng trong phạm vi bán kính 3 dặm tăng lên 1 USD thì số lượt người được nhà hàng phục vụ trong một năm sẽ tăng lên khoảng 2,3399 lượt về mặt trung bình.