

Chương Trình Giảng Dạy Kinh tế Fulbright

Học kỳ Thu năm 2014

Các Phương Pháp Phân Tích Định Lượng

BÀI TẬP 7

HỒI QUY TUYẾN TÍNH ĐƠN BIẾN (II)

Ngày Phát: Thứ ba 11/11/2014

Ngày Nộp: Thứ ba 02/12/2014

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

Bản điện tử gửi lên <http://www.fetp.edu.vn/vn/tai-nguyen/hoc-vien-hien-tai/>

Câu 1: (40đ)

Dữ liệu bao gồm diện tích canh tác và sản lượng lúa thu được của 436 hộ dân thuộc nhóm nghèo nhất nước được ghi nhận có dạng như sau:

(*) DỮ LIỆU 1			(*) DỮ LIỆU 2		
STT	Dientich (m2)	Sanluong (kg)	STT	Dientich (ha)	Sanluong (tấn)
1	1000	363	1	0.1	0.36
2	1008	500	2	0.1	0.50
3	1010	500	3	0.1	0.50
...
...
436	28000	9378	436	2.8	9.38

(*) Dữ liệu 2 bản chất là dữ liệu 1, chỉ chuyển đổi đơn vị diện tích và sản lượng.

Sử dụng DỮ LIỆU 2 và dùng công cụ Regression trong Excel để ước lượng kết quả hàm hồi quy mẫu, cho kết quả như Bảng sau đây:

SUMMARY OUTPUT DỮ LIỆU 2						
<i>Regression Statistics</i>						
Multiple R	0.8077					
R Square	0.6524					
Adjusted R Square	0.6516					
Standard Error	0.9281					
Observations	436					
<i>ANOVA</i>						
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>	
Regression	1	701.7099	701.7099	Value 6	1.2278E-101	
Residual	434	373.8702	0.8615			
Total	435	1075.5801				
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	Value 1	0.0600	3.3652	0.0008	0.0839	0.3196
dientich (ha)	2.7698	0.0970	Value 2	Value 3	Value 4	Value 5

- a. Anh/chị hãy tính giá trị còn thiếu Value1 trong bảng trên và viết phương trình hàm hồi quy mẫu từ kết quả tính toán được? Giải thích ý nghĩa của hệ số hồi quy $\hat{\beta}_2$?
- b. Tính giá trị còn thiếu Value2? Theo anh/chị, diện tích đất canh tác có ảnh hưởng (một cách có ý nghĩa thống kê ở mức ý nghĩa $\alpha = 5\%$) đến sản lượng lúa hay không? Giải thích bằng cách đưa ra các giả thuyết hợp lý và kiểm định giả thuyết đó?
- c. Anh/chị hãy tính toán xác suất mắc sai lầm loại I mà anh/chị sẽ gặp phải bằng bao nhiêu nếu anh chị bác bỏ giả thuyết H_0 ở câu b?
- d. Với mức ý nghĩa như trên, anh/chị hãy tính toán giá trị còn thiếu Value4 và Value5? Anh chị có nhận xét gì từ kết quả tính toán được?
- e. Anh/chị hãy tính toán giá trị còn thiếu Value6. Anh chị có nhận xét gì từ kết quả tính toán. Giải thích bằng cách đưa ra giả thuyết hợp lý và kiểm định giả thuyết đó.
- f. Nếu một hộ dân có diện tích đất canh tác là 02ha, anh/chị hãy cho biết sản lượng lúa trung bình của hộ dân này là bao nhiêu?
- g. Bằng kinh nghiệm thực tiễn, một hộ dân trong quá trình khảo sát cho rằng “*khi diện tích đất canh tác tăng thêm 01 ha, thì trung bình sản lượng lúa sẽ tăng lên ít nhất 03 tấn lúa*”. Anh/chị hãy kiểm định lại nhận định trên.
- h. Nếu ta thay dữ liệu đầu vào từ DỮ LIỆU 2 thành DỮ LIỆU 1 và chạy lại mô hình hồi quy như trên. Anh chị hãy lập luận để có được phương trình hàm hồi quy mẫu mới./.

Anh/chị sử dụng file dữ liệu microcredit.dta sau đây để thực hành phân STATA tiếp theo.

STT	Tên biến	Mô tả biến
1	nh	Mã hộ
2	year	Năm quan sát
3	villid	Mã làng (xã) (1, 2, 3, 4)
4	thanaid	Mã thôn (bản) (1, 2, 3, 4)
5	agehead	Tuổi chủ hộ (tuổi)
6	sexhead	Giới tính chủ hộ (Nam=1, Nữ=0)
7	educhead	Số năm đi học chủ hộ (năm)
8	famsize	Quy mô hộ (người/hộ)
9	hhland	Diện tích hộ (decimal); 01 decimal = 40.46m ²
10	hhasset	Tài sản hộ (Tk)
11	expfd	Chi tiêu lượng thực hộ (Tk/năm)
12	expnfd	Chi tiêu ngoài lượng thực hộ (Tk/năm)
13	exptot	Chi tiêu bình quân đầu người của hộ (Tk/năm)
14	dmmfd	Hộ tham gia chương trình tín dụng vi mô dành cho nữ (Yes=1, No=0)
15	dfmfd	Hộ tham gia chương trình tín dụng vi mô dành cho nam (Yes=1, No=0)
16	weight	Trọng số hộ
17	vaccess	Xã có đường giao thông kết nối (Yes=1, No=0)
18	pcirr	Tỷ lệ diện tích được tưới tiêu (%)
19	rice	Giá gạo (Tk/kg)
20	wheat	Giá lúa mì (Tk/kg)
21	milk	Giá sữa (Tk/lít)
22	potato	Giá khoai tây (Tk/kg)
23	egg	Giá trứng (Tk/04 quả)
24	oil	Giá dầu ăn (Tk/kg)

Câu 2: (5đ)

Aznh/chị hãy xóa ngẫu nhiên 20 quan sát (không được xóa các quan sát liền kề nhau, hay cách nhau một cách có hệ thống). Ghi chú lại các quan sát đã xóa và lưu lại file dữ liệu với tên mới là tên của anh chị theo dạng tên đầy đủ không dấu, họ và chữ lót viết tắt, không khoảng trắng.

Ví dụ: Tên của bạn là **Bùi Quốc An**, thì lưu tên file là: **anbq.dta**

Anh/chị sử dụng file này để trả lời các câu hỏi tiếp theo.

Anh chị vẫn trả lời các câu hỏi trên file word như các bài tập trước; file điện tử đính kèm sẽ bao gồm Do-files và File dữ liệu mới của anh chị sau khi đã xóa ngẫu nhiên 20 hộ quan sát.

Câu 3: (35đ)

- Sau khi đã xóa 20 quan sát, anh/chị hãy tạo 01 bảng thống kê mô tả cho các biến định lượng sau (**agehead, educhead, famsize, hhland, pcirr**) với một số tham số thống kê đặc trưng như sau (**số quan sát, giá trị nhỏ nhất, giá trị lớn nhất, trung bình, trung vị, độ lệch chuẩn**). Nhận xét kết quả tính toán được.
- Tạo bảng tần suất cho các biến định tính (không bao gồm biến **nh** và biến **year**). Nhận xét kết quả tính toán được.
- Anh/chị hãy vẽ đồ thị hiển thị tròn (pie graph) thể hiện tỷ lệ % số hộ có tham gia chương trình tín dụng vi mô dành cho nữ và các hộ không tham gia chương trình tín dụng vi mô dành cho nữ.
- Hãy vẽ đồ thị cột (column graph) biểu diễn mức chi tiêu trung bình đầu người của các hộ có tham gia chương trình tín dụng vi mô dành cho nữ và các hộ không tham gia chương trình tín dụng vi mô dành cho nữ. Theo anh chị, mức chi tiêu của 2 nhóm trên có khác nhau một cách có ý nghĩa thống kê ở mức $\alpha=5\%$ hay không?
- Anh/chị hãy tính hệ số tương quan giữa biến **exptot** với từng biến: **expfd** và **expnfd**. Theo anh chị, trong 2 biến trên, biến nào có mối tương quan chặt hơn với biến **exptot**? Vẽ đồ thị scatter graph thể hiện 2 mối liên hệ trên.

Câu 4: (20đ)

- Anh chị hãy tính hệ số tương quan giữa biến **exptot** và biến **hhland**? Kết luận từ kết quả có được?
- Nếu định nghĩa hộ có diện tích lớn hơn 50 decimal là hộ gia đình có diện tích lớn và hộ có diện tích ≤ 50 decimal là hộ có diện tích nhỏ. Theo anh/chị, chi tiêu bình quân của 2 nhóm hộ này có khác nhau một cách có ý nghĩa ở mức $\alpha = 5\%$ hay không?
- Anh chị hãy chạy phương trình hồi quy đa biến giữa biến **exptot** và các biến **eduhead, agehead, farmsize, hhland, pcirr**. Theo anh chị các biến nào có tác động lên biến **exptot** một cách có ý nghĩa thống kê? Các biến nào không có tác động lên biến **exptot** một cách có ý nghĩa thống kê ở mức $\alpha = 5\%$.
- Một hộ gia đình có tuổi của chủ hộ là 45 tuổi, trình độ học vấn của chủ hộ là 10 năm, nhà có 5 thành viên, diện tích 2000 decimal, diện tích được tưới tiêu là 50%. Anh/chị hãy dự báo trung bình chi tiêu đầu người của hộ đó là bao nhiêu? Khoảng tin cậy 95% sự biến thiên của chi tiêu trung bình đầu người dự báo là bao nhiêu?

-----HẾT-----