

Chương Trình Giảng Dạy Kinh tế Fulbright

Học kỳ Thu năm 2015

Các Phương Pháp Phân Tích Định Lượng

BÀI TẬP 11

BIẾN PHỤ THUỘC ĐỊNH TÍNH

Ngày Phát: Thứ ba 22/12/2015

Ngày Nộp: Thứ ba 29/12/2015

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

Bản điện tử nộp trước 8h20 tại địa chỉ: <http://www.fetp.edu.vn/vn/tai-nguyen/hoc-vien-hien-tai/>

Lưu ý:

- Học viên làm bài tập 11 theo Nhóm đề án môn học. Tuy nhiên, mỗi thành viên đều phải viết cho mình đáp án riêng theo cách hiểu của mỗi cá nhân. Tên của tất cả các thành viên trong Nhóm phải được ghi rõ trong bài (phần cuối cùng của bài làm).*
- Tất cả các file liên quan đến bài tập 11 (gồm file dữ liệu, log-file, do-file,...) phải được nộp về cho Ban Giảng viên như các bài tập trước.*

1. Giới thiệu tình huống

Đây là một bài tập sử dụng tình huống nghiên cứu thực tế. Học viên cần thiết phải làm quen với việc sử dụng phần mềm STATA và bộ dữ liệu tiếp cận nguồn lực của hộ gia đình (VARHS) để thiết kế mô hình, chọn lọc biến, và diễn giải mô hình.

Trong bài tập này học viên sẽ xây dựng mô hình đánh giá các nhân tố ảnh hưởng đến xác suất tham gia các hoạt động phi nông nghiệp của chủ hộ gia đình ở nông thôn. Học viên cần đọc bảng câu hỏi điều tra từ bộ dữ liệu VARHS 2012 (*xem file **Questionnaire.xls** đính kèm*) và phải đối chiếu bảng câu hỏi với các biến tương ứng trong các file dữ liệu STATA được phát (có tất cả 4 file: Q1.dta; Q5.dta; Q7D.dta và Q10A.dta).

2. Giải thích biến

Biến phụ thuộc:

Mục 5 (Employment status) với 5 biến phụ thuộc cần dùng, và nằm trong tệp dữ liệu Q5.dta.

- | | |
|--|---------|
| • Việc làm được trả công: | p24q1a_ |
| • Sản xuất nông/lâm/thủy sản: | p24q1b_ |
| • Phi nông, mua bán, cung cấp dịch vụ: | p24q1c_ |
| • Hải lượm săn bắn: | p24q1d_ |
| • Làm việc nhà: | p24q1e_ |

Biến giải thích:

-  Nhân khẩu học, Giáo dục, Tình trạng sức khỏe

Sheet 1A-Roster1 và Sheet 1A-Roster2 trong bảng câu hỏi (*file Questionnaire.xls*) và các biến tương ứng trong file STATA đính kèm (*tệp dữ liệu Q1.dta*).

- Giới tính: p1q3_
- Năm sinh: p1q4_
- Tình trạng sức khỏe: p1q5_
- Hôn nhân: p1q8_
- Giáo dục: p1q10_

Tài sản

Sheet 7D-Asset trong bảng câu hỏi (*file Questionnaire.xls*) và các biến tương ứng trong file STATA đính kèm (*tệp dữ liệu Q7D.dta*).

- Loại tài sản: p38ma_
- Số lượng: p38q1_
- Giá trị hiện tại: p38q3_

Vốn xã hội (tham gia đoàn thể tổ chức chính trị)

Sheet 10A-Groups1 trong bảng câu hỏi (*file Questionnaire.xls*) và các biến tương ứng trong file STATA đính kèm (*tệp dữ liệu Q10A.dta*).

- Loại hình tổ chức tham gia: p42q1_

3. Yêu cầu bài làm

Bước 1. Đọc dữ liệu, làm sạch, và chuẩn bị dữ liệu

Đối với biến phụ thuộc là trạng thái làm việc của chủ hộ:

- Sắp xếp dữ liệu theo vị trí địa lý và số thứ tự thành viên của hộ gia đình (tỉnh/quận/xã/hộ/số thứ tự thành viên).
- Tạo các biến giả hay dummy variable: **wage**, **production**, **service**, **hunt**, **housework** (tương ứng với các biến p24q1a_ / p24q1b_ / p24q1c_ / p24q1d_ / p24q1e_ /) nhận giá trị bằng 1 nếu thành viên hộ gia đình trả lời là có, và nhận giá trị bằng 0 nếu hộ gia đình trả lời là không.
- Giữ lại các quan sát là chủ hộ, còn lại bỏ.
- Bỏ các biến không cần thiết trong tệp dữ liệu ban đầu.
- Lưu tệp dữ liệu với tên mới là **depvar.dta**.

Đối với các biến giải thích về nhân khẩu học của chủ hộ:

- Xếp dữ liệu theo vị trí địa lý và số thứ tự thành viên của hộ gia đình (tỉnh/quận/xã/hộ/số thứ tự thành viên).
- Tạo biến **hhsz** mô tả số thành viên trong hộ gia đình (sử dụng lệnh `egen`).
- Tạo biến giả **gender** với giá trị bằng 1 cho thành viên là nam, và 0 nếu là nữ.

- Tạo biến giả **illness** với giá trị bằng 1 cho người có bệnh, 0 nếu không. Lưu ý bỏ những quan sát có giá trị 98 (missing value).
- Tạo biến giả **marriage** = 1 cho người hiện có gia đình, 0 nếu không. **Lưu ý bỏ những quan sát có giá trị 98 (missing value).**
- Tạo biến **age** là độ tuổi. **Lưu ý bỏ những quan sát có giá trị 99 (missing value).**
- Tạo biến **educ** là số năm đi học. **Lưu ý bỏ những quan sát có giá trị 98 (missing value).**
- Giữ lại các quan sát là chủ hộ, còn lại bỏ.
- Loại bỏ những biến không cần thiết.
- Lưu tệp dữ liệu với tên **demo.dta**.

Đối với các biến giải thích về tài sản của hộ:

- Xếp dữ liệu theo vị trí địa lý và số thứ tự thành viên của hộ gia đình (tỉnh/quận/xã/hộ/số thứ tự thành viên).
- Tạo biến **value** là giá trị từng loại tài sản (**bảng giá hiện tại * số lượng**)
- Tính tổng giá trị tài sản của từng hộ bằng lệnh collapse.
- Lưu tệp dữ liệu với tên mới là **asset.dta**.

Đối với các biến giải thích vốn xã hội của hộ:

- Xếp dữ liệu theo vị trí địa lý và số thứ tự thành viên của hộ gia đình (tỉnh/quận/xã/hộ/số thứ tự thành viên).
- Tạo biến giả **commy** nhận giá trị bằng 1 nếu thành viên hộ gia đình là Đảng viên, nhận giá trị 0 nếu không phải.
- Tính tổng số thành viên trong hộ là Đảng viên bằng lệnh collapse.
- Lưu lại tệp dữ liệu với tên mới là **commy.dta**.

Bước 2. Hợp nhất các tệp dữ liệu

- Hợp các bộ dữ liệu đã tạo ra bằng mã tỉnh/quận/xã/mã hộ, bắt đầu bằng **depvar**, **demo**, **asset**, và **commy**.
- Lưu bộ dữ liệu với tên **all.dta**.

Bước 3. Kiểm tra dữ liệu trước khi chạy mô hình hồi quy

- Trong bộ dữ liệu đã hợp nhất có 3126 quan sát.

Bước 4. Thống kê mô tả

- Lập bảng thống kê mô tả dữ liệu, thực hiện các kiểm định cần thiết, giải thích ý nghĩa của các kiểm định. Học viên có thể cần sử dụng biểu đồ minh họa. Ước lượng ma trận tương quan và giải thích?

Bước 5. Xây dựng mô hình hồi quy và ước lượng

- Mô hình xác suất tham gia vào một trong các hoạt động phi nông được xây dựng như sau:

$$P(I_i=1|X_i)=G[\beta_0+\beta_1X_1+\dots+\beta_kX_k+u_i]$$

Trong đó:

- $X_1, \dots, X_k$ là các biến giải thích được lọc từ các bước trên.
 - G là hàm xác suất tích lũy (Cumulative Distribution Function – CDF)
- a. Chạy mô hình hồi quy xác suất tuyến tính (LPM – Linear Probability Models) với các biến phụ thuộc lần lượt là **wage, production, service, hunt, housework**, và biến giải thích là **hhszise, gender, illness, marriage, age, educ, value, commy** (*Lưu ý: sử dụng lựa chọn robust để điều chỉnh phương sai thay đổi trong mô hình*).
- Ước lượng khả năng dự báo của mô hình đối với biến phụ thuộc là làm công ăn lương (**wage**). Anh/chị hãy kiểm định xem cả hai biến giải thích là vốn tài sản và vốn xã hội có đồng thời không có ý nghĩa thống kê trong mô hình này hay không?
- b. Chạy mô hình hồi quy Logit với lần lượt năm biến phụ thuộc trên.
- Ước lượng tác động biên tại các giá trị trung bình của các biến giải thích?
- Ước lượng khả năng dự báo của mô hình đối với biến phụ thuộc là làm công ăn lương (**wage**). Anh/chị hãy kiểm định xem cả hai biến giải thích là vốn tài sản và vốn xã hội có đồng thời không có ý nghĩa thống kê trong mô hình này hay không?
- c. Lặp lại Câu b, nhưng sử dụng mô hình hồi quy Probit.

Bước 6. Giải thích kết quả tính toán ở từng câu

- Giải thích ngắn gọn ý nghĩa của kết quả, gồm cả ý nghĩa thống kê và ý nghĩa kinh tế.
- So sánh giữa các mô hình LPM/Logit/Probit.

---HẾT---