

Trường Chính sách công và Quản lý Fulbright

Học kỳ Hè 2018

Đánh giá chính sách

BÀI TẬP 1

Ngày Phát: Thứ sáu 06/07/2018

Ngày Nộp: Thứ sáu 13/07/2018

Bản in nộp trước 8h20 tại Hộp nộp bài tập

Bản điện tử nộp trước 8h20 tại địa chỉ: <https://www.fsppm.fuv.edu.vn/vn/dang-nhap/>

Câu 1:

Xem xét 03 (ba) câu hỏi về quan hệ nhân quả sau:

- Nhiều doanh nghiệp, đặc biệt là ở phía Nam, là doanh nghiệp nhỏ, và sở hữu tư nhân thuộc gia đình. Có phải doanh nghiệp gia đình chậm phát triển hơn các doanh nghiệp có cấu trúc sở hữu đa dạng?
- Tác động của việc học kinh tế học là gì thay vì học lịch sử đối với mức lương của sinh viên đại học?
- Tác động của lãi suất cho vay mua nhà lên số dự án xây dựng nhà mới là gì?

Đối với mỗi câu hỏi, hãy trả lời các vấn đề sau:

- a. Biến kết quả là gì và biến can thiệp là gì?
- b. Xác định phản thực của biến kết quả Y_{0i} và Y_{1i} ?
- c. Kênh truyền dẫn tác động trực tiếp từ biến can thiệp đến biến kết quả là gì?
- d. Có những nguyên nhân nào gây ra tác động chệch lựa chọn (selection bias) khi so sánh đơn thuần kết quả theo các hình thức can thiệp khác nhau? Nhận định về hướng chệch và giải thích tại sao?

Câu 2:

Với những câu hỏi sau đây, chúng ta sử dụng bộ dữ liệu **bm.dta** từ một thử nghiệm ngẫu nhiên được thực hiện bởi Marianne Bertrand và Sendhil Mullainathan. Các nhà nghiên cứu gửi 4870 bộ hồ sơ nguyện tạo đến các nhà tuyển dụng theo các thông báo tuyển dụng ở Boston và Chicago năm 2001. Các hồ sơ khác nhau về tên ứng viên, và việc lựa hồ sơ được tiến hành ngẫu nhiên cho các vị trí đang tuyển dụng. Một số tên gợi ý cho nhà tuyển dụng biết ứng cử viên là người da trắng, và một số tên đặc trưng cho người da đen. Các nhà nghiên cứu muốn tìm hiểu liệu những hồ sơ mà tên của ứng viên hàm ý họ là người da đen thì sẽ nhận được ít số cuộc gọi hẹn phỏng vấn hơn là các tên hàm ý người da trắng?

- a. Bộ dữ liệu có hai biến giả gồm *female* và *computerskill*, trong đó quan sát nhận giá trị 1 cho biết ứng viên là nữ (*female*), và tương tự là ứng có kỹ năng máy tính. Hãy lập các bảng chéo các biến này theo biến *black* (1-ứng viên có tên hàm ý họ là người da đen). Sử dụng lệnh tham khảo dưới đây, từ kết quả hãy cho biết tình trạng giới tính và kỹ năng máy tính có cân bằng giữa các nhóm phân theo chủng tộc hay không?

tab female black, col

- b. Lập các bảng tương tự cho các chỉ tiêu khác bao gồm: giáo dục (education) và số lượng công việc đã từng làm trước (ofjobs). Các biến này lần lượt có 5 và 7 giá trị khác nhau. Từ kết quả, hãy cho biết giáo dục và số lượng công việc trước đây có cân bằng giữa các nhóm chủng tộc không?
- c. Sử dụng lệnh **summarize** để xem xét giá trị trung bình và độ lệch chuẩn cho biến số năm kinh nghiệm làm việc (yearsexp) cho từng nhóm người da đen và da trắng (sử dụng tùy chọn **if**). Bình luận sự cân bằng về số năm kinh nghiệm giữa các chủng tộc?
- d. Bạn nhận định như thế nào từ các kết quả liên quan đến đặc tính của hồ sơ ứng viên ở các câu trước? Tại sao chúng ta lại quan tâm liệu các đặc tính đó có cân bằng giữa các chủng tộc?
- e. Biến cần quan tâm trong tập dữ liệu là biến cuộc gọi (**call**, giá trị 1 cho biết ứng viên đã nhận được cuộc gọi từ nhà tuyển dụng). Bạn có thấy sự khác biệt về tỷ lệ cuộc gọi lại theo chủng tộc không?
- f. Bạn kết luận gì từ kết quả thử nghiệm của Bertand và Mullainathan về tính thiên vị liên quan đến chủng tộc trong tuyển dụng?

Câu 3:

Trong phần trước chúng ta đã nghiên cứu thử nghiệm Bertand và Mullainathan về dữ liệu hồ sơ ứng viên. Trong câu hỏi này, sử dụng bộ dữ liệu cps.dta thu được từ điều tra dân số hiện tại hàng tháng (CPS) năm 2001 (một cuộc điều tra thị trường lao động lớn). Bộ dữ liệu này có 8891 cá nhân tham gia, sống ở Boston và Chicago. Chúng ta muốn sử dụng bộ dữ liệu này để so sánh kỹ năng của các cá nhân da trắng và da đen, kết quả tuyển dụng, và sự khác biệt với các kết quả phát hiện từ Câu hỏi 2.

- a. Tập dữ liệu chứa biến giáo dục education với bốn giá trị (1 - học sinh bỏ học thời trung học; 2- học sinh tốt nghiệp trung học; 3 - cao đẳng; 4 - đại học và cao hơn). Sử dụng biến giáo dục để tạo một biến giả mới cho những hồ sơ mà ứng viên có bằng cao đẳng trở lên (tức là có nhóm có bằng cao đẳng, nhóm có bằng đại học và cao hơn). Tỷ lệ người có trình độ giáo dục từ cao đẳng trở lên là bao nhiêu?
- b. Thực hiện **kiểm định t** để kiểm tra liệu tỷ lệ hồ sơ có bằng từ cao đẳng trở lên của các nhóm người da trắng và da đen có bằng nhau không? Học viên có thể sử dụng lệnh **ttest**. Ví dụ, nếu biến cần kiểm tra là somecol thì học viên viết lệnh như sau:

```
ttest somecol, by(black)
```

Báo cáo:

1. Tỷ lệ hồ sơ có bằng từ cao đẳng trở lên cho người da trắng?
2. Tỷ lệ hồ sơ có bằng từ cao đẳng trở lên cho người da đen?
3. Sự khác biệt tỷ lệ giữa hai hai nhóm sắc tộc?
4. Trị kiểm định t?
5. Giá trị p-value đối với giả thuyết không (tỷ lệ có bằng cao đẳng trở lên giữa hai nhóm sắc tộc là cân bằng)
6. Bạn có phát hiện ra bằng chứng cho thấy sự khác biệt về tỷ lệ có bằng cao đẳng trở lên đối với người da trắng và da đen trong bộ dữ liệu CPS từ các kết quả trên?
- c. Tính trị kiểm định t cho kiểm định về giá trị trung bình của số năm làm việc (yearsexp) là bằng nhau giữa hai nhóm sắc tộc. Bạn có phát hiện ra bằng chứng cho thấy số năm làm việc là khác nhau đối với người da trắng và da đen?
- d. Thảo luận kết quả từ phần (b) và (c). Tại sao kết luận của bạn cho các biến giáo dục và số năm kinh nghiệm lại khác nhau? Tại sao chúng ta lại quan tâm liệu các biến đó giống nhau giữa hai chủng tộc?
- e. Tính trị kiểm định t đối với kiểm định tỷ lệ người có việc làm hay không giữa hai

nhóm sắc tộc (**employed**) là bằng nhau. Bạn có phát hiện bằng chứng cho thấy tỷ lệ này có sự khác biệt lớn giữa nhóm da trắng và da đen?

- f. Theo các kết quả từ phần (d) và (e), bạn có thể kết luận gì về phân biệt chủng tộc trong tuyển dụng lao động thông qua bộ dữ liệu CPS?
- g. Các kết quả này có liên hệ như thế nào với kết quả thu được từ Câu hỏi 2?

---HẾT---