

NHỮNG ỨNG DỤNG CỦA THỐNG KÊ TRONG NGHIÊN CỨU

NGUYỄN MINH TUẤN(*)

Trong công tác nghiên cứu khoa học ở nhà trường, các tác giả thường sử dụng phương pháp thu thập thông tin bằng cách sử dụng bảng điều tra như là một công cụ nghiên cứu. Một trong những nguyên nhân chính của cách thu thập này là có thể thu thập thông tin một cách nhanh chóng trong một thời gian tương đối ngắn. Những đối tượng mà đề tài nghiên cứu nhắm đến sẽ được gửi bảng câu hỏi và những câu trả lời thu thập được sẽ làm cơ sở cho những phân tích xa hơn. Những dữ liệu thu được này có thể thuộc loại định tính như: tốt, xấu, trung bình... hay thuộc loại định lượng như các con số. Thông thường các tác giả chỉ dừng lại việc mô tả kết quả thống kê được mà thiếu đi những suy luận xa hơn. Điều này thật là đáng tiếc nếu chúng ta biết rằng trong kiến thức về thống kê có hai phần thống kê mô tả và thống kê suy diễn, trong đó thống kê suy diễn chiếm phần lớn kiến thức và từ đó chúng ta có thể rút ra rất nhiều kết luận từ những mẫu điều tra được. Những kết luận này có thể dùng làm cơ sở cho những quyết định của chúng ta.

Sau đây chúng ta hãy xem xét những kết luận thông thường từ một số cuộc nghiên cứu và xem chúng ta có thể đi xa như thế nào với công cụ thống kê:

Ví dụ 1: Đối với dữ liệu định tính.

Trong một cuộc thăm dò 1200 sinh viên tại trường đại học, sinh viên được yêu cầu đánh

giá chất lượng giảng dạy tại trường có được nâng cao hơn so với khi họ bắt đầu vào trường hay không. Kết quả thăm dò cho thấy có 30% sinh viên đánh giá chất lượng giảng dạy xấu hơn; 28% đánh giá tốt hơn; còn lại 42% đánh giá không thay đổi.

Một số tác giả gặp phải trường hợp này đã kết luận rằng chất lượng giảng dạy đã xuống cấp do 30% lớn hơn 28% và phải có nhiều nỗ lực hơn nữa để thúc đẩy chất lượng giảng dạy. Dĩ nhiên, việc nâng cao chất lượng giảng dạy là một điều cần thiết và cần được làm thường xuyên liên tục nhưng việc đánh giá chất lượng giáo dục đã xuống cấp thì hơi vội vàng. Một ví dụ để phản bác điều này là trường hợp bạn tung đồng tiền 58 lần. Kết quả mặt sấp xuất hiện 30 lần và mặt ngửa xuất hiện 28 lần thì bạn không thể kết luận rằng xác suất mặt sấp xuất hiện sẽ lớn hơn mặt ngửa. Thực tế, chúng ta biết rằng nếu tung đồng tiền càng nhiều lần thì các xác suất này sẽ càng xấp xỉ bằng nhau và bằng 0.5.

Với việc áp dụng thống kê suy diễn cho dữ liệu định tính chúng ta sẽ giải quyết vấn đề như sau:

(*) Giáo viên Cơ hữu Khoa Kinh tế & Quản trị Kinh doanh Đại Học Mở Bán công TP.HCM

Ký hiệu p là tỷ lệ sinh viên nhận thấy chất lượng giảng dạy xấu đi.

Các giả thuyết: $H_0: p \leq 0.5$

$H_a: p > 0.5$.

Kích cỡ mẫu dùng để phân tích

$$n = 30\% * 1200 + 28\% * 1200 = 696.$$

Số sinh viên trả lời chất lượng

xấu đi sẽ tuân theo phân phối chuẩn

với $\mu = 0.5 * n = 0.5 * 696 = 348$

$$\sigma = \text{SQRT}(0.25 * n) = \text{SQRT}(0.25 * 696) = 13.19.$$

$$x = 0.30 * 1200 = 360$$

$$\text{Giá trị thống kê kiểm định } z = \frac{x - \mu}{\sigma} = (360 - 348) / 13.19 = 0.91.$$

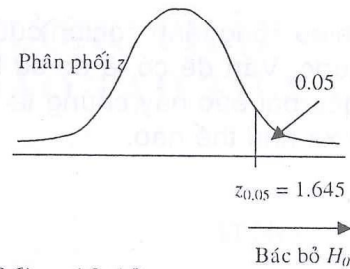
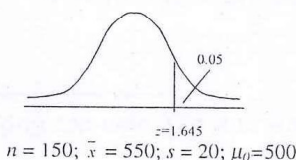
Do 0.91 nhỏ hơn 1.645 nên ta không bác bỏ H_0 . Tức là không thể kết luận đa số sinh viên đánh giá chất lượng giảng dạy của trường đã giảm đi.

Ở đây chúng ta nhận thấy rằng nếu có sự khác biệt "đáng kể" giữa số sinh viên đánh giá chất lượng nâng cao hay giảm đi thì chúng ta mới có thể kết luận chất lượng đã tăng lên hay giảm xuống.

Ví dụ 2: Với dữ liệu định lượng

Trong một cuộc điều tra nghiên cứu trước đây sinh viên ở TPHCM chi tiêu bình quân 500 (ngàn) đồng một tháng để trang trải cho các chi phí như ăn ở, tài liệu học tập... Do giá cả tăng lên trong thời gian gần đây sinh viên phàn nàn rằng chi phí này không đủ trang trải cuộc sống hàng ngày nếu họ không đi làm thêm vào cuối tuần. Để quyết định có nên tăng học bổng hay không chính quyền quyết định tiến hành điều tra. Mẫu 150 học sinh được lựa chọn ngẫu nhiên và kết quả chi tiêu trung bình là 550 (ngàn) một tháng với độ lệch chuẩn là 20 (ngàn). Hãy quyết định xem có đủ bằng chứng chi phí đã tăng lên hay không?

Qua ví dụ 1 ở trên chúng ta sẽ không vội vã để kết luận rằng do 550 (ngàn) lớn hơn 500 (ngàn) nên chi phí đã tăng lên. Chúng ta sẽ sử dụng thống kê suy diễn để kết luận về trường hợp này.



Giả thuyết: $H_0: \mu \leq 500$

$H_a: \mu > 500$

$$z = \frac{\bar{x} - \mu_0}{s / \sqrt{n}} = \frac{550 - 500}{22 / \sqrt{150}} = 27.835$$

Do 27.835 lớn hơn 1.645 nên ta bác bỏ H_0 với độ tin cậy 95%

Kết luận: Chi phí đã tăng lên.

Trên đây là hai ví dụ về việc áp dụng thống kê để có thể kết luận vấn đề một cách sâu sắc hơn. Còn rất nhiều vấn đề khác trong nhà trường có thể sử dụng công cụ thống kê để nghiên cứu như:

- Nghiên cứu số sinh viên tốt nghiệp/rớt giữa các khoa, từ đó xem khoa sinh viên lựa chọn theo học có ảnh hưởng đến tỷ lệ tốt nghiệp của sinh viên hay không.

- Nghiên cứu lãnh vực lựa chọn nghề nghiệp của sinh viên mới ra trường: dịch vụ; công nghiệp; nghiên cứu khoa học... có phụ thuộc vào giới tính của sinh viên hay không.

- Xây dựng mô hình quan hệ giữa kết quả học tập với giới tính, điểm thi vào trường, thời gian học tập tại nhà...

- Nghiên cứu chiều cao của sinh viên trong trường để xem chiều cao của sinh viên có tăng lên trong 5 năm gần đây hay không.

TÓM TẮT:

Hiện tại, nhiều công trình nghiên cứu tại trường chúng ta đã không khai thác hết những dữ liệu đã thu thập được. Vấn đề có lẽ do sự hạn chế trong việc nhận định vấn đề dưới khía cạnh thống kê. Thông qua bài báo này chúng ta khảo sát một vài đề tài để xem với công cụ thống kê chúng ta có thể đi xa như thế nào.

SUMMARY:

At present, many research projects at our university have not fully exploited the data gathered. The problem is likely due to the inadequacy of looking at the matter from a statistics viewpoint. In this article, we shall examine some projects to see how further we can go with the aid of statistics tools.