

SỬ DỤNG THỐNG KÊ VÀ QUAN NIỆM SAI LẦM VỀ THỐNG KÊ của người sử dụng thông tin thống kê

TS. J.V.Bhalerao

Trợ lý giáo sư, Viện MGV quản lý và nghiên cứu, Nashik, Ấn Độ

Giới thiệu

Từ “Thống kê” có thể được sử dụng trong hình thức số nhiều hoặc số ít. Khi sử dụng ở dạng số ít, nó đề cập đến các nhánh của kiến thức về trao đổi giữa các nguyên tắc, các thiết bị hoặc phương pháp thu thập, phân tích và diễn giải dữ liệu. Nếu ở dạng số nhiều, thống kê có nghĩa là thu thập dữ liệu trong hệ thống với mục đích xác định đối tượng.

Từ Thống kê có nguồn gốc từ tiếng Latin là “Status” hoặc Ý “Statista” hoặc Đức “Statistik” mỗi từ đó đều có ý nghĩa là hành chính nhà nước, cụm từ số liệu thống kê đã được sử dụng để hiển thị số liệu và sự kiện trong những lĩnh vực khác nhau như kinh tế, nhân khẩu học và các tình huống chính trị của một quốc gia, thống kê thường được đưa ra thông qua chính quyền địa phương trong suốt những thập kỷ qua.

Vào thế kỷ 19 từ khi số liệu thống kê phát triển vượt bậc như một nhánh của khoa học và cố gắng đi sâu vào mọi lĩnh vực hoạt động của con người. Các câu hỏi cơ bản trong sử dụng số liệu thống kê là “**Thống kê là một ngành khoa học hay nghệ thuật?**”. Trong từ điển ý nghĩa từ của khoa học

là một sự tích lũy kiến thức được chấp nhận, được hệ thống hóa và xây dựng với những điều khoản tham chiếu, để khám phá ra bản chất hoặc hoạt động của luật pháp nói chung, còn từ nghệ thuật là một kỹ năng thực hiện hành động nào đó theo thói quen bằng kinh nghiệm, học tập hoặc quan sát. Nếu cả hai định nghĩa được nghiên cứu cẩn thận người ta có thể kết luận rằng số liệu thống kê là cả nghệ thuật cũng như khoa học. Thống kê là khoa học vì phương pháp về cơ bản có hệ thống và có các ứng dụng, thống kê cũng là nghệ thuật vì ứng dụng thành công của thống kê phụ thuộc vào kỹ năng và kinh nghiệm đặc biệt của các nhà thống kê ở một mức độ nào đó.

Trong 100 năm qua việc sử dụng các phương pháp thống kê như vậy đã mở rộng, mà ngày nay hầu như bất kỳ nhánh kiến thức nào cũng sử dụng các phương pháp này. Có thể nói rằng tất cả các khoa học sẽ khiếm khuyết nếu không có số liệu thống kê. Một mình nó có thể không có khả năng truyền tải thông tin ý nghĩa, nhưng một số ngành cùng sử dụng số liệu thống kê có thể truyền đạt ý nghĩa cho nhiều ngành khoa học khác, có thể nói rằng “**Khoa học không số liệu thống kê không**

có quả; thông kê không có khoa học không có gốc rễ". Những khái niệm về nhánh khoa học, nghiên cứu hoặc quản lý được xây dựng, thử nghiệm hoặc thay đổi với sự giúp đỡ của thông kê.

Sử dụng số liệu thông kê

Quan sát cho thấy rằng ứng dụng và sử dụng số liệu thông kê đã phát triển đáng kể ở ba lĩnh vực chính: khoa học, kinh doanh và chính phủ. Chính phủ đã sử dụng số liệu thông kê từ thời kỳ hoàng đế Akbar, nhưng ứng dụng số liệu thông kê về sau này được tìm thấy trong các lĩnh vực khoa học và kinh doanh. Những phát minh khác nhau và thay đổi công nghệ trong khoa học dẫn đến cuộc cách mạng công nghiệp. Đó là thời gian từ khi doanh nghiệp bắt đầu trở nên phức tạp vì nhiều quy định của chính phủ, tăng thuế, công nghệ thay đổi, vấn đề lao động, và nguồn lực quản lý; một trong số đó là nhân lực và sau cùng là cắt giảm chi phí để tối đa hóa lợi nhuận.

Để hiểu được tình hình của tất cả những vấn đề này và để giữ cho doanh nghiệp bền vững thì sử dụng số liệu thông kê là cần thiết. Ngoài ra việc phân tích và giải thích tình hình xã hội cũng làm tăng nhu cầu về thông kê, để đánh giá điểm mạnh và điểm yếu của nền kinh tế thì chính phủ cần luôn thu thập và phân tích thông tin, tất cả các ngành kinh tế như tiêu thụ, sản xuất, phân phối, trao đổi, vv... được mô tả, tương quan cũng nhờ sự giúp đỡ của thông kê. Do đó, tất cả những yếu tố trên cũng làm tăng nhu cầu về sử dụng thông kê.

Nói chung, có thể quan sát thấy được sự tăng trưởng qua số liệu thông kê, sử dụng các phương pháp thông kê khác nhau trong các lĩnh vực đã thúc đẩy sự phát triển của khoa học thông kê. Các kỹ thuật thông kê đã được cải tiến đến một mức giúp cho chi phí, thời gian thu thập và phân tích dữ liệu

giảm xuống nhờ sử dụng công nghệ máy tính và số hóa. Trong tương lai khoa học thông kê sẽ tiếp tục cải thiện và việc sử dụng số liệu thông kê có cấu trúc hơn.

(1). Thông kê đơn giản hoá những dấu tích của dữ liệu: Việc tổ chức và trình bày dữ liệu như là một cách để họ nhìn nhận ra ý nghĩa. Tại thời điểm bất kỳ, ảnh, sơ đồ hoặc trình bày đồ họa của dữ liệu làm giảm độ phức tạp, tăng sức hấp dẫn và tạo điều kiện cho việc so sánh thông tin dữ liệu, ngoài ra còn giúp cho việc hiểu biết liên quan tới các vấn đề khác.

(2). Thông kê trình bày sự kiện trong hình thức nhất định: Các dữ liệu thu thập được trình bày trong các hình thức chữ số để có độ chính xác trong khi truyền đạt thông tin có ý nghĩa. Một tuyên bố nói rằng không khí lạnh đang tăng không phải là một thông kê, nhưng khi người ta nói rằng nhiệt độ đã tăng 20°C so với ngày trước đó thì là số liệu thông kê, bởi vì cho thấy chính xác nhiệt độ lạnh đã tăng lên.

(3). Thông kê là cơ sở so sánh: Việc sử dụng tối đa số liệu thông kê là để so sánh và tìm ra nguyên nhân một mối quan hệ hiệu quả nêu thực sự tồn tại mối quan hệ đó. Việc so sánh giúp tìm kiếm sự thay đổi và sự thay đổi này có thể xảy ra trong tương lai. Ví dụ hiệu ứng Nhà xanh, sự suy giảm của tầng ozone, sự nóng lên toàn cầu là những ví dụ mà số liệu thông kê dự báo chính xác những thay đổi, cũng gợi ý những biện pháp khắc phục những gì phát hiện ra và lợi ích của những biện pháp đó.

(4). Dự báo: Đây là chức năng quan trọng của thông kê dùng để dự đoán về mối quan hệ các sự kiện trong tương lai. Công cụ mạnh mẽ nhất cho các doanh nghiệp và cả nền kinh tế.

(5). Thống kê giúp xây dựng và thử nghiệm giả thuyết: Thống kê giúp xây dựng và thử nghiệm giả thuyết ngoài ra nó giúp đưa ra quyết định hợp lý. Dựa trên việc hiểu biết các biện pháp, mô tả những tác động của hành động và phản ứng, thực hiện mối tương quan giữa các sự kiện độc lập để có được thông tin về một bức tranh đầy đủ của vấn đề.

Khoa học thống kê đã cung cấp lợi ích cho nhân loại, người sử dụng số liệu thống kê cũng nên là các chuyên gia. Đây là một con dao hai lưỡi, nếu sử dụng đúng thì có thể chứng minh là điều có ích, nhưng nếu lạm dụng sẽ có thể là thảm họa. Số liệu thống kê giúp người sử dụng trình bày sự thật như họ đang có với yêu cầu có kiến thức phù hợp về phương pháp thống kê, điều tra kỹ lưỡng, thu thập và phân tích dữ liệu để được đảm bảo. Có những hạn chế nhất định về thống kê do nhiều người thống kê sử dụng sai. Còn lại một phần của bài viết là việc xem qua các quan điểm sai lầm về Thống kê hoặc lạm dụng số liệu thống kê.

Quan niệm sai lầm trong thống kê:

Hiện nay nghiên cứu ngày càng nhiều hơn và có ý nghĩa hơn trong mọi lĩnh vực, vì vậy càng thiết thực với lĩnh vực quản lý. Khi mà người chưa qua đào tạo sử dụng số liệu thống kê để rút ra kết luận nó có thể gây hiểu lầm, những con số không có lỗi mà do người sử dụng.

(1). Nghiên cứu phải có thử nghiệm cho giả thuyết: Mỗi nghiên cứu đưa ra và được trình bày trong luận án phải có một kiểm định tham số hoặc phi tham số phù hợp của giả thuyết Mis (Mis là hệ thống quản lý thông tin) của nhà nghiên cứu. Nếu các dữ liệu thu thập cho nghiên cứu có thể không phải trình bày bằng số, nó sẽ không thể áp dụng thử nghiệm cho giả thuyết. Trong thực tế, hạn chế lớn nhất của thống kê là cần dữ liệu được trình bày bằng

số, không thể phân tích dữ liệu định tính như sử dụng số liệu thống kê trực tiếp.

(2). Sử dụng phương pháp tỷ lệ phần trăm: Phương pháp tỷ lệ phần trăm có thể không được sử dụng để xác nhận kết quả bởi vì đây là phương pháp rất đơn giản, thực tế chứng minh rằng mức độ thay đổi thường thể hiện bằng tỷ lệ phần trăm. Ngoài ra, còn dùng chuyển đổi các dữ liệu vào cơ sở dữ liệu chung dùng để so sánh, nếu ai đó sử dụng phương pháp tỷ lệ phần trăm để so sánh thì đây là phương pháp dễ dàng và hợp lý.

(3). Lấy mẫu phải được thực hiện: Cho dù việc lấy mẫu phải được thực hiện hay không điều đó phụ thuộc vào quy mô tổng thể được xem xét. Giả sử rằng một bộ phận kiểm soát chất lượng của một công ty sản xuất có 11 mục cần kiểm tra chất lượng và 120 mặt hàng từ một nhà cung cấp để kiểm tra chất lượng. Chất lượng của nó được quyết định nếu: 10% các mặt hàng được kiểm tra ngẫu nhiên đạt yêu cầu. Làm thế nào để giám sát chất lượng đưa ra quyết định? Câu trả lời là lựa chọn một trong 11 mục ở trên, và 12 mặt hàng từ 120 mặt hàng để kiểm tra, nếu họ sản xuất theo tiêu chuẩn chất lượng thì sẽ được chấp nhận. Nghe có vẻ hợp lý?. Điều đó là tốt, do đó đối với một quy mô tổng thể nhỏ 11 mục tại sao chỉ lấy một mẫu? Khi người ta có thể có được kết quả chính xác 100% độ tin cậy bằng cách thực hiện điều tra toàn bộ tổng thể và cũng không liên quan đến nhiều nguồn dữ liệu thì tại sao phải sử dụng mẫu? Điều đó có thể là vì chuyên môn kém và sự thiếu cẩn thận của một số người sử dụng số liệu thống kê.

(4). Phương pháp lấy mẫu phải luôn luôn là ngẫu nhiên không?: Tất cả các số liệu thống kê theo thời gian nhấn mạnh vào việc không phụ thuộc các yếu tố bên ngoài cả thủ tục cũng như kết

quả. Nếu một nhà nghiên cứu đang làm một cuộc khảo sát thăm dò ý kiến liên quan đến một số vấn đề hiện tại. Nếu cứ cố đi để lấy mẫu ngẫu nhiên có thể không nhận được dữ liệu đầy đủ. Đơn giản là phải lấy mẫu ngẫu nhiên cập nhật danh sách của người dân và sau đó là lựa chọn ngẫu nhiên người trả lời. Làm thế nào có thể có được một danh sách như vậy và sau đó chọn mẫu? Ngược lại để có thể có được một danh sách như vậy, người lấy mẫu thuận tiện và gọi nó là lấy mẫu như ngẫu nhiên. Ví dụ như một cuộc khảo sát các bác sĩ liên quan đến nhi khoa là một mẫu là được lựa chọn, thay vì các nhà nghiên cứu gặp bác sĩ ngẫu nhiên thì họ chọn gặp bác sĩ nhi khoa để thu thập thông tin và họ nói rằng họ đã chọn mẫu ngẫu nhiên.

(5). Ứng dụng không chắc chắn làm cho số liệu thống kê không tin cậy: Giả sử số liệu thống kê được tìm thấy từ quá khứ cho thấy là 20% các hoạt động phẫu thuật nhất định bởi một bác sĩ là thất bại. Và biết rằng 4 ca phẫu thuật cuối cùng của bác sĩ đã thành công, bạn sẽ đoán bệnh nhân thứ 5 cho hoạt động phẫu thuật này? Xác suất thất bại là 20% điều đó không có nghĩa là trong năm trường hợp người ta phải có một bệnh nhân bị thất bại. Do Luật Thống kê không chính xác, nên trong năm trường hợp được xem xét, tất cả chúng có thể thành công. Bởi bản báo cáo duy nhất có ý nghĩa là mẫu lớn hơn và phạm vi rộng hơn để kết quả ở trên mức trung bình.

(6). Phân tích thống kê cần có kiểm định một mẫu nhỏ hay mẫu lớn hoặc kiểm định chi-bình phương: Ngày nay, phân tích thống kê bằng máy tính có sẵn cho các nhà nghiên cứu. Vì vậy, để thấy rằng phân tích được thực hiện và thống kê được sử dụng thì các nhà nghiên cứu cố gắng sử dụng bất kỳ các thử nghiệm trên đó, vì mang lại giá trị cho các

đề tài nghiên cứu hoặc luận văn. Điều đó thực sự cần thiết? Có liên quan đến các mục tiêu nghiên cứu? Đây là những câu hỏi không ai quan tâm. Như ví dụ sau một câu hỏi đã có và các câu hỏi khác liên quan đến ý thích và ghét hương vị khác nhau của các loại kem. Các nhà nghiên cứu cũng đã thu thập dữ liệu liên quan đến các biến nhân khẩu học một trong số đó là thu nhập của một người. Áp dụng kiểm định chi-bình phương giữa nhóm thu nhập của một người và ý thích hương vị của một loại kem. Kết luận các nhà nghiên cứu là nhóm có thu nhập cao là những người thích hương vị bơ trộn với đường. Tại sao vậy? Điều quan trọng là phải hiểu rằng các thử nghiệm không thể áp dụng cho bất kỳ dữ liệu được thu thập, thứ hai là mặc dù kiểm định chi-bình phương cho thấy mối quan hệ đó là tương quan nhưng không thật để người dùng sử dụng nó như là một phép suy luận.

(7). Cỡ mẫu có thể khác nhau cho mỗi câu hỏi trong bảng câu hỏi: Nếu một bảng câu hỏi chứa 10 câu hỏi. Cỡ mẫu cho phân tích của mỗi câu hỏi phải giống nhau. Nếu kích thước mẫu quyết định là 200 người và giả sử rằng 150 người đã trả lời tất cả các câu hỏi trong khi còn lại 50 người có trả lời vài câu hỏi trong bảng câu hỏi. Cỡ mẫu có liên quan sẽ là 150 người. Có thể câu trả lời của 50 người còn lại không liên quan đến câu hỏi trong bảng câu hỏi. Do đó 50 người trả lời này phải được xử lý như phần dư thừa và nên được bỏ qua ở phân tích sau cùng. Vì vậy, để đảm bảo quy định, nhà nghiên cứu cỡ mẫu phải đảm bảo để có được đầy đủ câu trả lời của người được hỏi.

(8). Thu thập dữ liệu quan trọng hơn mục tiêu hoàn thiện: Các nhà nghiên cứu luôn luôn nghĩ thu thập dữ liệu là quan trọng hơn mục tiêu hoàn thiện. Mục tiêu có thể được xây dựng sau để phù

hợp với phân tích. Thống kê không phải là lừa bịp mà họ có thể chỉnh sửa hay thay đổi các dữ liệu để phù hợp với sự mong đợi các nhà nghiên cứu. Giả sử nhà nghiên cứu thu thập các thông tin về doanh số bán hàng của tất cả các loại thức ăn như đồ ăn nhanh, mặt hàng thực phẩm như ngũ cốc và dầu ăn vv... được sử dụng để nấu thức ăn. Mục tiêu hoàn thành là mô hình tiêu thụ loại thức ăn đồ ăn nhanh trên thị trường. Việc sử dụng tất cả các dữ liệu thu thập còn lại và các nguồn lực được sử dụng cho những bộ dữ liệu của mình đó là những gì? Nếu mục tiêu là để tìm thương hiệu nào phổ biến hơn. Khi nhận ra rằng thương hiệu phổ biến thì thông tin không được thu thập đầy đủ. Làm thế nào đối với những dữ liệu thu thập được? Thống kê có thể giúp bạn trong việc này.

(9). Nghiên cứu nên mang lại kết quả rõ ràng và mong muốn: Các câu hỏi thường xuyên là yêu cầu thống kê thực hiện những gì để có được kết quả rõ ràng? Thống kê luôn luôn cố gắng đưa ra sự thật khách quan. Khi một nghiên cứu thực hiện tại sao các nhà nghiên cứu luôn luôn nên đưa ra kết quả rõ ràng? Ngay cả khi nó đã chứng minh rằng các dữ liệu thu được cho thấy kết quả không có ích, nó cũng giúp xây dựng vấn đề nghiên cứu của họ trong tương lai, và tiết kiệm thời gian. Có thể mục đích duy nhất là đánh giá tài liệu có liên quan.

(10). Thuyết phục mọi người sử dụng số liệu thống kê: Có người đi qua các quảng cáo mà nói rằng 80% số người tin chắc rằng sản phẩm của chúng tôi là tốt nhất? Thật vậy, nếu 8 trong số 10 người đang đánh giá các sản phẩm là tốt nhất, bạn

sẽ xử lý kết quả lựa chọn là tốt nhất? Không bao giờ kích thước mẫu thật được trích dẫn trong quảng cáo như vậy. Nhưng ít nhất mỗi lần mua sản phẩm cũng là một lần suy nghĩ. Tất cả chúng ta nói chuyện về TRP (chương trình nghiên cứu công nghệ) trên truyền hình. Làm thế nào nhiều người hiểu được cách tính? Các thiết bị được cố định vào các máy truyền hình đếm số lần bạn chuyển sang một kênh đặc biệt vào cùng một thời gian. Giả sử bạn bắt đầu xem truyền hình với kênh X lúc 9 giờ đêm, nó ghi nhận bạn như là một người xem kênh đó cho chương trình 9 giờ, ngay cả khi bạn không xem các chương trình từ đầu tới khi kết thúc, chuyển sang kênh khác ngay lập tức hoặc sau đó một thời gian.

Kết luận

Theo giáo sư Chou "báo cáo này đã được lấy cảm hứng từ việc lạm dụng số liệu thống kê trong phần thực hành của các học viên không được đào tạo phương pháp thống kê, người dùng không hiểu rõ kết quả thống kê, người dùng không trung thực sử dụng số liệu thống kê dẫn đến đánh lừa và gây nhầm lẫn hơn là sự rõ ràng". Có thể nói rằng lạm dụng số liệu thống kê vì thiếu kiến thức kỹ thuật, sử dụng cách kiểm tra không phù hợp hoặc thậm chí dữ liệu không phù hợp, ý kiến cá nhân và nhận thức của các nhà điều tra thường là không nhớ những hạn chế trong khi phân tích và giải thích dữ liệu là vấn đề. Lớn nhất là mất lòng tin về số liệu thống kê: ***có ba loại dối trá - "Lies" có nghĩa là nói dối vô hại, "Damn lies" là nói dối dẫn đến hậu quả đáng kể và "statistics" là thống kê.*** Đây là nghiên cứu và những người làm thống kê phải làm rõ điều này.

Công Hoan (dịch)

Nguồn bài viết: www.researchersworld.com/vol2/PAPER_08.pdf