

NĂNG LỰC PHÂN TÍCH DỮ LIỆU TRONG NGHIÊN CỨU KHOA HỌC GIÁO DỤC CỦA GIẢNG VIÊN ĐẠI HỌC

Trần Trung

Học viện Dân tộc

Email: trung1978@gmail.com

Article history

Received: 12/11/2024

Accepted: 30/12/2024

Published: 05/02/2025

Keywords

Educational science, data analysis skills, applied statistics, statistical analysis

ABSTRACT

University lecturers' research in the field of educational sciences is broadly described as specific and systematic. High-quality educational sciences research can only be achieved with logical and modern research methods in the context of global industrialization, modernization and international integration. The use of statistical theories and data analysis in educational scientific research designs is increasingly emphasized, particularly quantitative research methods. This article analyzes the data analysis capacity in educational scientific research of university lecturers through researching, synthesizing and analyzing relevant documents. The results show that applied statistics help with analyzing data from previous studies on teaching and learning effectiveness. The study is the basis for proposing some solutions to improve the lecturers' required knowledge of appropriate statistical terminology and conventions to present research results, skills of interpreting and presenting statistical analysis clearly and accurately, while evaluating the research context and its limitations.

1. Mở đầu

Việc sử dụng công nghệ, các kỹ thuật phân tích dữ liệu, kết hợp với phương pháp nghiên cứu phù hợp là định hướng của hoạt động KH-CN hiện đại, trong đó có lĩnh vực khoa học giáo dục (KHGD) trong giai đoạn hiện nay. Trong những năm qua, KHGD đã đóng vai trò góp ý, phản biện và định hình chính sách, dự báo xu thế phát triển của giáo dục trong tương lai, giới thiệu và kiểm chứng các mô hình, phương pháp, và lý thuyết mới trong KHGD, đồng thời đóng vai trò kết nối hệ thống giáo dục Việt Nam với quốc tế (Nguyễn Tiến Trung và cộng sự, 2024). Nghiên cứu trong lĩnh vực KHGD là một hoạt động nghiên cứu khoa học đặc thù và có tính hệ thống của giảng viên ở trường đại học. Để tạo ra các nghiên cứu có chất lượng cao, các nghiên cứu KHGD phải áp dụng những phương pháp nghiên cứu logic, hiện đại nhằm đáp ứng yêu cầu thời kỳ CNH, HĐH và hội nhập quốc tế. Việc sử dụng các lý thuyết thống kê và phân tích dữ liệu trong các thiết kế nghiên cứu KHGD ngày càng được chú trọng, đặc biệt phương pháp nghiên cứu định lượng hiện đang được áp dụng rộng rãi trong các công trình nghiên cứu KHGD. Nghiên cứu của Hoàng Văn Hào và Phạm Hoàng Điệp (2020) cũng cho thấy, kỹ năng phân tích dữ liệu được xem là một phần quan trọng trong năng lực nghiên cứu khoa học của giảng viên.

Việc thiếu nền tảng về thống kê ứng dụng và phân tích dữ liệu không chỉ cản trở các nhà nghiên cứu KHGD trong việc công bố bài báo trên các tạp chí khoa học uy tín, mà còn ảnh hưởng đến khả năng khai thác các dữ liệu sẵn có trong đời sống, nhất là trong thời đại dữ liệu lớn, trí tuệ nhân tạo và sự phát triển của công nghệ thông tin và truyền thông và mạng xã hội như hiện nay. Kruppa và cộng sự (2021) đã chỉ ra sự khác biệt giữa giảng viên dạy về thống kê và sinh viên trong các lĩnh vực ngoài thống kê (bao gồm khoa học xã hội) trong việc hiểu về các khái niệm thống kê. Điều này dẫn đến việc thiếu hiệu quả trong việc dạy và học thống kê đối với các sinh viên không chuyên (về thống kê). Trong một khảo sát gần đây với các nghiên cứu trong KHGD của Việt Nam, các tác giả đã thống kê chỉ có 22.4% số công bố là sử dụng phương pháp định lượng, thấp hơn hẳn so với con số tương ứng sử dụng phương pháp định tính là 50.2%. Nguyên nhân của hiện tượng này có thể là bởi các nhà nghiên cứu KHGD tại Việt Nam chưa được đào tạo nhiều về phương pháp nghiên cứu định lượng trong suốt quá trình đào tạo từ đại học, thạc sĩ cho đến tiến sĩ (Vu et al., 2023).

Bài báo đề xuất cấu trúc năng lực phân tích dữ liệu trong nghiên cứu KHGD cho giảng viên đại học và trình bày các năng lực phân tích dữ liệu trong nghiên cứu KHGD của giảng viên đại học thông qua phương pháp nghiên cứu, tổng hợp và phân tích các tài liệu, văn bản.

2. Kết quả nghiên cứu

2.1. Tổng quan nghiên cứu về phân tích dữ liệu trong khoa học giáo dục

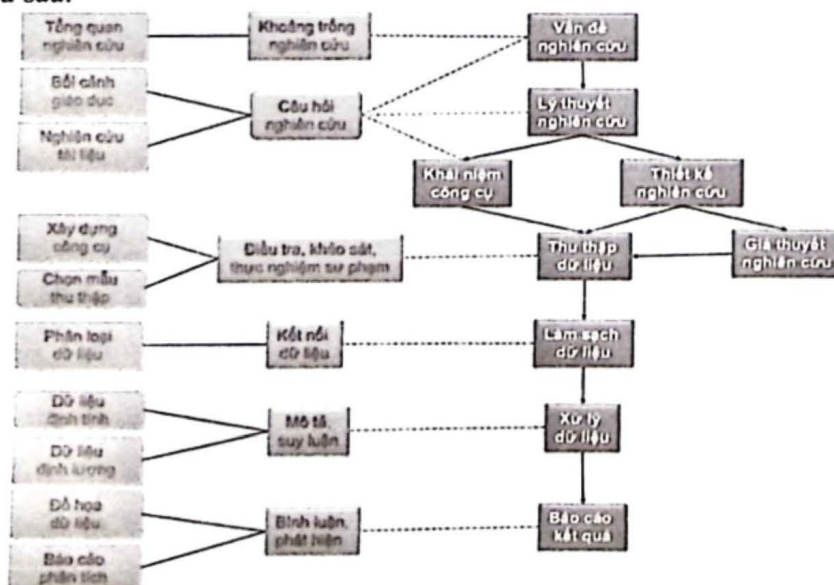
Bakker và Wicherts (2014) nhấn mạnh tầm quan trọng của việc các nhà nghiên cứu phải có hiểu biết vững chắc về các phương pháp thống kê và các giả thuyết cơ bản của chúng. Họ nhấn mạnh tầm quan trọng của việc phân tích dữ liệu một cách minh bạch và theo hướng có thể tái tạo lại để đảm bảo tính hợp lệ và độ tin cậy của kết quả nghiên cứu. Trong những năm đầu của thế kỉ XXI, nhiều trường học đã bắt đầu tạo ra trải nghiệm học tập được cá nhân hóa thông qua việc sử dụng công nghệ trong lớp học. Nhờ vào sự tiến bộ của lượng dữ liệu thu thập được, các kĩ thuật học máy đã được áp dụng để cải thiện chất lượng giáo dục bao gồm các lĩnh vực liên quan đến học tập và phân tích nội dung (Lai et al., 2014). Việc sử dụng các kĩ thuật này cho mục đích giáo dục là một lĩnh vực đầy hứa hẹn nhằm phát triển các phương pháp khám phá dữ liệu từ các cơ sở giáo dục có tính toán và khám phá các mẫu có ý nghĩa.

Để giải quyết những thách thức của phân tích dữ liệu, các nhà nghiên cứu đã đề xuất nhiều giải pháp và khuyến nghị khác nhau. Nghiên cứu của Gigerenzer (2018) đã lập luận cho việc thúc đẩy năng lực thống kê và năng lực đánh giá rủi ro cho các nhà nghiên cứu cũng như cho các đối tượng rộng hơn. Họ đề xuất dạy tư duy thống kê và kỹ năng ra quyết định để cải thiện việc diễn giải và truyền đạt các phát hiện dựa trên dữ liệu, nhấn mạnh tầm quan trọng của việc tích hợp kỹ năng tính toán và lập trình vào chương trình đào tạo khoa học xã hội. Những kỹ năng này có thể giúp các nhà nghiên cứu quản lý, xử lý và phân tích hiệu quả các tập dữ liệu lớn và phức tạp, đang trở nên phổ biến ngày càng nhiều trong nghiên cứu khoa học xã hội. Các nhà nghiên cứu cũng xem xét các sai lệch tiềm tàng, các yếu tố văn hóa và bối cảnh, cũng như sử dụng phù hợp các cách tiếp cận phân tích định tính và định lượng. Tổ chức các khóa bồi dưỡng, phát triển chuyên môn và thúc đẩy năng lực thống kê là các yếu tố thiết yếu để nâng cao kỹ năng phân tích dữ liệu và đảm bảo tính hợp lệ và độ tin cậy của kết quả nghiên cứu trong các lĩnh vực khoa học xã hội.

Ở Việt Nam hiện nay, bên cạnh các kỹ thuật phân tích thống kê cơ bản như thống kê mô tả, khoảng tin cậy, kiểm định giả thuyết và phân tích phương sai, các phương pháp phân tích dữ liệu sau cũng được các nhà nghiên cứu trong lĩnh vực khoa học xã hội thường xuyên áp dụng, bao gồm phân tích hồi quy, phân tích hồi quy logistic, phân tích cụm (cluster analysis), phân tích đa biến, và phân tích dữ liệu chuỗi thời gian. Cụ thể, phân tích hồi quy thường được dùng trong Xã hội học và Nhân khẩu học để dự đoán các yếu tố ảnh hưởng đến các biến số như thu nhập, giáo dục, hoặc sự hài lòng với cuộc sống; trong giáo dục học và tâm lý học, phân tích hồi quy thường được áp dụng để dự đoán kết quả học tập của học sinh dựa trên các yếu tố như điểm số trước đó, lượng thời gian học tập, và các biến số tâm lý khác. Phân tích phân cụm được sử dụng để nhóm các quốc gia hoặc khu vực dân cư có các đặc điểm chung về mặt KT-XH, văn hóa. Phân tích nhân tố được sử dụng để xác định các mẫu ẩn trong dữ liệu bằng cách nhìn vào mối quan hệ giữa các biến. Trong giáo dục học và tâm lý học, phân tích nhân tố có thể được sử dụng để phân tích cấu trúc của các kỹ năng học tập, hoặc để tìm ra các yếu tố tâm lý ẩn có thể ảnh hưởng đến hành vi của con người.

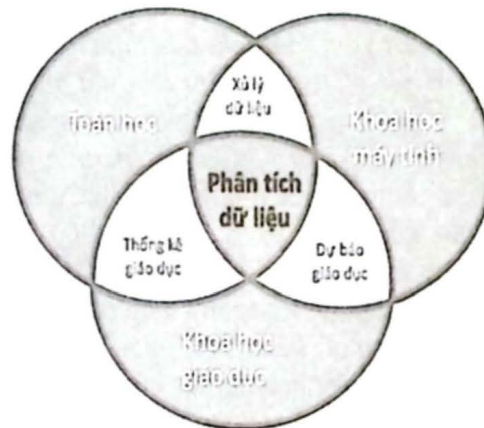
2.2. Cấu trúc năng lực phân tích dữ liệu trong nghiên cứu khoa học giáo dục cho giảng viên đại học

Căn cứ đặc trưng của hoạt động giáo dục, sư phạm tương tác, tác giả xác định lược đồ phân tích dữ liệu trong nghiên cứu KHGD như sau:



Hình 1. Lược đồ phân tích dữ liệu trong nghiên cứu KHGD

Dựa vào kết quả nghiên cứu của Nguyễn Thị Mỹ Lộc và cộng sự (2023), tác giả đề xuất cấu trúc năng lực phân tích dữ liệu trong nghiên cứu KHGD của giảng viên đại học bao gồm các thành tố sau:



Hình 2. Cấu trúc năng lực phân tích dữ liệu trong nghiên cứu KHGD

Có thể mô tả các thành tố năng lực cụ thể như sau:

- **Năng lực xác định luồng dữ liệu trong thiết kế nghiên cứu:** Thiết kế nghiên cứu là bản kế hoạch, là lộ trình cho toàn bộ nghiên cứu. Nó sẽ chỉ rõ một số phần quan trọng nhất của nghiên cứu như những người tham gia, cách thức, thời gian và địa điểm dữ liệu sẽ được thu thập và tất nhiên là phân tích. Nghiên cứu sẽ được triển khai dựa trên kế hoạch cụ thể này. Thiết kế nghiên cứu KHGD có thể chia làm hai nhóm: Thiết kế nghiên cứu quan sát (Observational Study) gồm thiết kế nghiên cứu mô tả (Discriptive study): *Thiết kế nghiên cứu trường hợp (Case series)*; *thiết kế nghiên cứu cắt ngang (Cross-sectional study)*; *thiết kế nghiên cứu tương quan (Correlational study)*. Thiết kế nghiên cứu phân tích (Analytical study): *Thiết kế nghiên cứu hồi nguyên (Case-control study)*; *thiết kế nghiên cứu thuần tập (Cohort study)*. Thiết kế nghiên cứu can thiệp (Experimental study) gồm thiết kế nghiên cứu can thiệp cộng đồng (Community Trials), thiết kế nghiên cứu can thiệp thực địa (Clinical Trials), thiết kế nghiên cứu can thiệp ngẫu nhiên (Randomized Trials): *Không có đối chứng (Randomized non-Controlled Trials)*; *có đối chứng (Randomized Controlled Trials - RCTs)*. Trong đó, giảng viên lĩnh vực KHGD cần làm chủ các thiết kế nghiên cứu và cách mà các thiết kế này được sử dụng trong quy trình triển khai các nghiên cứu về KHGD; với mỗi một dạng thiết kế nghiên cứu cần xác định rõ luồng dữ liệu để đảm bảo quy trình thực hiện trong KHGD.

- **Năng lực thu thập dữ liệu:** Thu thập dữ liệu được coi là một trong những công đoạn có ý nghĩa vô cùng quan trọng, ảnh hưởng trực tiếp đến kết quả nghiên cứu. Mặc dù vậy, việc thu thập dữ liệu cũng là công đoạn vất vả nhất, tốn nhiều thời gian, công sức và chi phí. Để việc thu thập dữ liệu hiệu quả, phục vụ đúng mục tiêu nghiên cứu, nhà nghiên cứu cần phải hiểu rất rõ về các loại dữ liệu mà nghiên cứu cần phải thu thập, đồng thời, phải nắm chắc quy trình và các phương pháp thu thập dữ liệu khác nhau tương ứng với từng loại dữ liệu khác nhau và phù hợp với các thiết kế nghiên cứu tương ứng. Trong lĩnh vực KHGD, các nghiên cứu cũng thường được thiết kế bao gồm: thiết kế nghiên cứu định lượng, thiết kế nghiên cứu định tính và các thiết kế nghiên cứu hỗn hợp. Thành tố năng lực này giúp nhà nghiên cứu có được một bộ dữ liệu chuẩn, đủ về số lượng và đảm bảo về chất lượng, phù hợp với mục tiêu nghiên cứu bằng hai quy trình và phương pháp thu thập dữ liệu cơ bản tương ứng với hai loại dữ liệu: sơ cấp và thứ cấp. Quy trình thu thập dữ liệu thứ cấp được triển khai với sáu bước cơ bản: xác định vấn đề nghiên cứu, xác định mục tiêu nghiên cứu, tra cứu nguồn dữ liệu, đánh giá độ tin cậy của nguồn dữ liệu, trích xuất dữ liệu, sắp xếp và trích dẫn dữ liệu. Đối với dữ liệu sơ cấp, quy trình và phương pháp thu thập chú trọng đến năm bước: xác định vấn đề nghiên cứu; chọn mẫu; thiết kế bảng hỏi; tổ chức điều tra, khảo sát bằng bảng hỏi; mã hoá và làm sạch dữ liệu. Mục tiêu cuối cùng của công cuộc thu thập dữ liệu là chất lượng của bộ dữ liệu thu thập được. Các phương pháp chọn mẫu (Sampling methods) gồm chọn mẫu phi xác suất ((Non-probability sampling): *Chọn mẫu thuận tiện (Convenient sampling)*; *Chọn mẫu chỉ tiêu (Quota sampling)*; *Chọn mẫu chủ đích (Purposive sampling)*; *Chọn mẫu hòn tuyết lăn (snowball sampling)* và chọn mẫu xác suất (probability sampling): *Chọn mẫu ngẫu nhiên đơn/ hệ thống (Simple/ Systematic random sampling)*; *Chọn mẫu xác suất tỉ lệ với kích cỡ quần thể (PPS sampling)*; *Chọn mẫu cụm (Cluster sampling)*; *Chọn mẫu ngẫu nhiên phân tầng (Stratified random sampling)*; *Chọn mẫu dây chuyền có kiểm soát (Respondent driven sampling)*. Ước tính cỡ mẫu (Sample size formula) gồm cỡ mẫu trong nghiên cứu định lượng (Quantitative research): *Ước tính theo chỉ tiêu thu thập; ước tính theo tham số thống kê của mô hình phân tích* và cỡ mẫu trong nghiên cứu định tính (Qualitative research): *Ước tính theo phương pháp quan sát, trao đổi, phỏng vấn*

sâu, thảo luận nhóm. Các loại dữ liệu cần thu thập gồm phân loại theo nguồn dữ liệu gồm *dữ liệu sơ cấp (Primary data)*, *dữ liệu thứ cấp (Secondary data)*; phân loại theo đặc điểm của dữ liệu gồm *dữ liệu định tính (Qualitative data)*, *dữ liệu định lượng (Quantitative data)*; phân loại theo thang đo của dữ liệu gồm *dữ liệu định danh (Nominal)*, *dữ liệu định khoảng (Interval)*, *dữ liệu định hạng (Ordinal)*, *dữ liệu định tỉ lệ (Ratio)*; phân loại theo biến phân tích gồm *dữ liệu liên tục (Continuous)*, *dữ liệu rời rạc (Discrete)*, *dữ liệu nhị phân (Binary)*, *dữ liệu phân nhóm (Nominal categorical)*, *dữ liệu số đếm (Count)*; phân loại theo thời gian thu thập dữ liệu gồm *dữ liệu cắt ngang (cross-sectional)*, *dữ liệu chuỗi thời gian (time series)*, *dữ liệu mảng (panel)*, *dữ liệu không gian (spatial)*, *dữ liệu gộp (meta)*. Thiết kế công cụ thu thập dữ liệu gồm công cụ thu thập dữ liệu định tính: *Phiếu quan sát, bảng hỏi cấu trúc, bảng hỏi bán cấu trúc, bảng hỏi phi cấu trúc*; công cụ thu thập dữ liệu định lượng: *Phiếu điều tra, khảo sát; data warehouse, data lake*. Thực hiện quy trình thu thập dữ liệu: Thu thập thử (Pilot); thu thập chính thức; đạo đức trong thu thập dữ liệu (Research Ethics); xử lý tình huống trong thu thập dữ liệu. Kết nối dữ liệu gồm kết nối dữ liệu tĩnh: *Kết nối trường dữ liệu (kiểu join)*, *kết nối dòng dữ liệu (kiểu union)* và kết nối dữ liệu động: *Kết nối thời gian thực (real-time)*, *kết nối dữ liệu gốc (API)*. Làm sạch dữ liệu nghiên cứu (Data cleaning) gồm kiểm định thang đo: *Hệ số tin cậy Cronbach's Alpha; Chỉ số Kappa; Độ tin cậy (Reliability); Độ giá trị (Validity); Độ nhạy (Sensitivity); Độ đặc hiệu (Specificity); ...* xử lý dữ liệu bất thường: *Giá trị khuyết (missing value); Giá trị đột biến (Out of range); Giá trị ngoại lai (Outlier)*. Do vậy, vấn đề quản lý dữ liệu, kiểm soát chất lượng dữ liệu và đảm bảo đạo đức nghiên cứu trong suốt quá trình thu thập dữ liệu là một trong những kỹ năng mà giảng viên KHGD để đảm bảo rằng, trước khi chuyển sang giai đoạn phân tích, xử lý dữ liệu, các nhà nghiên cứu về KHGD hoàn toàn yên tâm với một bộ dữ liệu thu về đáp ứng các tiêu chí: đúng, đủ và sạch. Thu thập dữ liệu đúng cách, dữ liệu thu về chuẩn xác và đúng mục đích sẽ là tiền đề quan trọng tạo nên thành công của nghiên cứu.

- *Năng lực phân tích dữ liệu định tính*: Phân tích dữ liệu định tính là khâu quan trọng và bắt buộc trong quy trình thực hiện một nghiên cứu định tính. Trước hết, giảng viên KHGD cần nắm vững các phương pháp thu thập dữ liệu định tính. Sau đó đảm bảo những nguyên tắc cơ bản trong phân tích dữ liệu định tính và quy trình phân tích dữ liệu định tính. Xử lý dữ liệu định tính gồm: xử lý dữ liệu kênh chữ (text); xử lý dữ liệu kênh tiếng (audio); xử lý dữ liệu kênh hình (video); phương pháp mã hóa dữ liệu định tính. Giảng viên KHGD cần thực hiện chặt chẽ từng bước trong quy trình phân tích dữ liệu định tính, sử dụng hiệu quả một số phần mềm chuyên dụng trong phân tích dữ liệu định tính, xác định tính hợp lệ và độ tin cậy của dữ liệu định tính và đạo đức nghiên cứu.

- *Năng lực phân tích thống kê mô tả*: Thống kê là khoa học và nghệ thuật thu thập, tóm tắt và phân tích dữ liệu có thể thay đổi ngẫu nhiên. Hai dạng tóm tắt và phân tích dữ liệu được dùng là thống kê mô tả và thống kê suy luận. Thống kê mô tả (descriptive statistics) là một phần quan trọng của các nghiên cứu về KHGD, là một trong những phương pháp xử lý dữ liệu được sử dụng khá phổ biến, được sử dụng để tóm tắt hoặc mô tả một tập hợp dữ liệu, một mẫu nghiên cứu dưới dạng số hay biểu đồ trực quan. Thông tin này giúp giảng viên KHGD thống kê dữ liệu thu thập được một cách khoa học, nhanh chóng và xử lý dữ liệu dễ dàng hơn. Thống kê suy luận (inferential statistics) là một phương pháp trong thống kê để đưa ra kết luận về một quần thể dựa trên một mẫu dữ liệu có sẵn. Thống kê suy luận giúp các nhà nghiên cứu đưa ra kết luận về toàn bộ quần thể dựa trên một mẫu dữ liệu hạn chế. Giảng viên KHGD cần tóm tắt dữ liệu bằng thống kê mô tả, các loại biến số và thang đo thường dùng trong nghiên cứu KHGD; thực hiện các phép phân tích thống kê cơ bản khi phân tích dữ liệu bằng thống kê mô tả như xác định tần suất, độ tập trung, độ phân tán của dữ liệu.

- *Năng lực phân tích thống kê suy luận*: Các kỹ thuật và mô hình thống kê suy luận được sử dụng để thực hiện khái quát về tổng thể dựa trên các mẫu. Nghĩa là với một mẫu nghiên cứu nhất định thống kê có thể giúp phân tích, mô tả và đưa ra suy luận, dự đoán về tổng thể nghiên cứu. Phân tích tương quan với các *hệ số tương quan Pearson; hệ số tương quan Spearman; hệ số tương quan Kendall tau-b (phi tham số)*. Phân phối và kiểm định Khi bình phương gồm *phân phối Khi bình phương; kiểm định Khi bình phương cho mô hình phù hợp nhất; kiểm định Khi bình phương cho bảng có số chiều 2x2 và 2x3*. Phân tích phương sai gồm *phân tích phương sai một yếu tố (ANOVA); phân tích phương sai hai yếu tố (ANCOVA); phân tích phương sai đa biến (MANOVA); phân tích phương sai phi tham số (Kiểm định Kruskal-Wallis cho hai biến số độc lập; Kiểm định Friedman cho hai biến số theo cặp)*. Phân tích hồi quy là kỹ thuật thống kê dùng để ước lượng phương trình phù hợp nhất với các tập hợp kết quả quan sát của biến phụ thuộc và biến độc lập; cho phép đạt được kết quả ước lượng tốt nhất về mối quan hệ chân thực giữa các biến số. Từ phương trình ước lượng được này, giảng viên KHGD có thể dự báo về biến phụ thuộc (chưa biết) dựa vào giá trị cho trước của biến độc lập (đã biết). Các mô hình hồi quy thường dùng bao gồm: *hồi quy tuyến tính (với biến liên tục, biến nhị phân, biến phân nhóm); hồi quy đa biến, hồi quy đa thức (vấn đề đa cộng tuyến, hoán chuyển dữ liệu,*

bootstrap); hồi quy logistic (một biến, đa biến, đa thức); hồi quy Poisson (với dữ liệu theo phân bố Poisson); hồi quy Cox (với dữ liệu thời gian; phân tích survival, phương pháp Kaplan-Meier, phương pháp Log rank test). Do đó, có một sự khác biệt lớn giữa thống kê mô tả và thống kê suy luận trong đó thống kê suy luận vẫn giữ vai trò mấu chốt trong các nghiên cứu thống kê. Thống kê suy luận liên quan đến các vấn đề ước tính các tham số của tổng thể và kiểm tra các giả thuyết. Giảng viên KHGD cần am hiểu một số lý thuyết cơ bản về thống kê suy luận và thực hiện được xử lý dữ liệu dựa trên thống kê suy luận. Giảng viên cần am hiểu các mô hình hồi quy thường sử dụng trong nghiên cứu KHGD trên để đánh giá mối quan hệ giữa biến phụ thuộc và biến độc lập, từ đó đưa ra những dự báo về mức độ tác động của các biến độc lập lên biến phụ thuộc trong mô hình.

- *Năng lực phân tích dữ liệu thực nghiệm sư phạm*: Nghiên cứu thực nghiệm trong KHGD được tổ chức dưới các hình thức can thiệp bởi các nhà nghiên cứu giáo dục đến các đối tượng nghiên cứu tìm hiểu về hiệu quả của một phương pháp giảng dạy, một mô hình giáo dục hay một chương trình đào tạo nào đó đối với đối tượng nghiên cứu. Thông thường, trong KHGD, nghiên cứu thực nghiệm được tiến hành theo phương pháp so sánh giữa nhóm thực nghiệm và nhóm kiểm soát (hoặc so sánh trước sau đối với cùng một nhóm sau thời gian tác động giáo dục). Nhóm thực nghiệm nhận được một tác động giáo dục (một phương pháp, phương tiện dạy học hoặc mô hình, giải pháp giáo dục mới, hoặc chương trình đào tạo mới), trong khi nhóm kiểm soát không chịu tác động của những mô hình, phương pháp, giải pháp mới mà chỉ chịu những tác động giáo dục hiện tại. Quá trình tác động giáo dục có thể diễn ra trong một thời gian dài hoặc thời gian nhất định do các nhà nghiên cứu đặt ra, những kết quả đo lường được so sánh giữa hai nhóm (hoặc một nhóm so với chính nó trước khi tác động) để xác định mô hình, giải pháp mới trong giáo dục có hiệu quả hơn hay không. Kiểm định giả thuyết khoa học trong thực nghiệm sư phạm dựa trên *biến số phụ thuộc và biến số độc lập; mức ý nghĩa, khoảng tin cậy, trị số p-value; kiểm định giả thuyết không; kiểm định t (t-test) với hai biến số độc lập; kiểm định t (t-test) với các biến số theo cặp; kiểm định Mann-Whitney cho hai biến số độc lập phi tham số; kiểm định Wilcoxon cho hai biến số theo cặp phi tham số*. Nghiên cứu thực nghiệm được coi là loại hình nghiên cứu mạnh nhất để đánh giá tác động của một vấn đề giáo dục mới đến quá trình giáo dục bởi nó cho phép các nhà nghiên cứu kiểm soát các biến khác nhau và chỉ tập trung vào tác động giáo dục cần quan tâm. Giảng viên KHGD cần có năng lực về thực hiện quy trình nghiên cứu thực nghiệm, sử dụng một số phương pháp thực nghiệm và thu thập, xử lý số liệu thực nghiệm trong KHGD.

- *Năng lực xử lý dữ liệu từ phân tích gộp*: Kỹ thuật phân tích gộp (meta-analysis) thực hiện thống kê các kết quả nghiên cứu của các kết quả nghiên cứu thuộc cùng một chủ đề nghiên cứu, hoặc thuộc các chủ đề nghiên cứu gần nhau. Giảng viên KHGD cần am hiểu phân tích gộp theo quy trình bảy bước và các kỹ thuật phân tích cơ bản nhất của một nghiên cứu phân tích gộp, ứng dụng trong lĩnh vực KHGD; cần phải có đủ kiến thức chuyên môn để có thể đánh giá được chất lượng của các tập dữ liệu trước khi quyết định sử dụng những kết quả của nghiên cứu phân tích gộp trong lĩnh vực KHGD.

- *Năng lực đồ họa và phiên giải dữ liệu*: Các biểu đồ thể hiện mối quan hệ dữ liệu và cung cấp tham số để phân tích trong nghiên cứu KHGD như: (1) *Các biểu đồ thể hiện mật độ dữ liệu*: bảng hai chiều (text tables), biểu đồ nhiệt (Heat maps, Highlight tables), bản đồ dữ liệu (symbol, Filled maps), biểu đồ bánh, vòng (pie chart, donut chart), biểu đồ thanh (bar plot, stacked bar, side-by-side bar), biểu đồ cây (treemaps), biểu đồ phân tán (circle views), biểu đồ hộp (box plot), biểu đồ dây (line plot, area charts), biểu đồ mạng nhện 360 (Spider map Diagram); (2) *Các biểu đồ thể hiện quan hệ dữ liệu*: biểu đồ phân bố (histogram plot), biểu đồ phân phối (scatter plot), biểu đồ hộp râu (box-whisker plot), biểu đồ so sánh điểm (Slope graphs), biểu đồ thác nước (Waterfall chart, Gantt bar), biểu đồ phân đoạn (Marimekko chart), tháp dân số (Chart population Pyramid), biểu đồ bullet, biểu đồ packed bubbles, biểu đồ sai số chuẩn (error bar plot), biểu đồ nguyên nhân tích lũy (Pareto chart), biểu đồ Radial, biểu đồ đa nhóm (Small Multiples), biểu đồ ti suất (Waffle charts), biểu đồ lợi nhuận (Lollipop chart), biểu đồ so sánh cùng kì (Dumbbell chart), biểu đồ hồi quy tuyến tính (Linear regression), biểu đồ hồi quy đa biến, biểu đồ hồi quy Poisson (dành cho dữ liệu số đếm), biểu đồ hồi quy logictis (dành cho dữ liệu nhị phân), biểu đồ phân tích dự báo (dành cho dữ liệu chuỗi thời gian); (3) *Các biểu đồ thể hiện biến đổi dữ liệu*: độ chênh (deviation), tương quan (Correlation), xếp hạng (Ranking), phân phối (Distribution), thời gian (Change over Time), tỉ trọng (Part-to-Whole), độ lớn (Magnitude), không gian (Spatial), lưu lượng (Flow).

Trong phân tích dữ liệu nghiên cứu KHGD, giảng viên cần hiểu các hàm suy luận trong phân tích dữ liệu là không chặt, vì vậy không có mô hình đúng mà chỉ có mô hình có ích; chỉ có kiểm định giả thuyết để bác bỏ giả thuyết sai. Dữ liệu chỉ cho biết thông tin, không cho biết kết quả, quan trọng là khả năng giải thích thông tin bằng kiến thức chuyên ngành KHGD. Đồng thời cần chú trọng khả năng tái lập kết quả nghiên cứu hơn là phát hiện mới.

3. Kết luận

Trong khoa học xã hội nói chung và KHGD nói riêng, thống kê ứng dụng giúp phân tích dữ liệu từ các cuộc khảo sát và nghiên cứu xã hội và có thể được sử dụng để xác định mối quan hệ giữa các biến xã hội, đánh giá tác động của các chính sách xã hội và phân tích dữ liệu từ các mô hình dự báo xã hội. Thống kê ứng dụng cũng có thể giúp tìm hiểu các xu hướng xã hội và đánh giá tác động của các yếu tố xã hội đến cuộc sống con người. Trong nghiên cứu khoa học xã hội trên thế giới hiện nay, các phương pháp truyền thống vẫn đóng vai trò quan trọng trong việc phân tích dữ liệu. Trong nghiên cứu KHGD, chọn mẫu, thống kê mô tả, kiểm định giả thuyết, phân tích phương sai và hồi quy là các công cụ cơ bản nhưng được sử dụng mạnh mẽ để khám phá mối quan hệ giữa các biến và đo lường tác động của chúng trong nghiên cứu (Levy & Lemeshow, 2013). Trong KHGD, giảng viên cần có kiến thức về thuật ngữ và quy ước thống kê phù hợp để trình bày kết quả nghiên cứu, kỹ năng báo cáo và diễn giải phân tích thống kê một cách rõ ràng và chính xác, đồng thời xem xét bối cảnh và các hạn chế của nghiên cứu. Thống kê ứng dụng giúp phân tích dữ liệu từ các nghiên cứu về hiệu quả giảng dạy và học tập, có thể được sử dụng để đánh giá tác động của các chương trình giáo dục, xác định mối quan hệ giữa các yếu tố giảng dạy và kết quả học tập, tìm hiểu các yếu tố ảnh hưởng đến việc học của người học. Thống kê ứng dụng cũng có thể giúp xác định các xu hướng, các mô hình trong học tập và đề xuất các phương pháp giảng dạy hiệu quả. Vì vậy cần xác định rõ các chỉ báo cụ thể cả các năng lực thành tố của năng lực phân tích dữ liệu trong nghiên cứu. khoa học xã hội để có cách thức tổ chức hoạt động phát triển năng lực này cho giảng viên lĩnh vực KHGD ở các trường đại học, trở thành người dẫn dắt cho các hoạt động nghiên cứu, nhóm nghiên cứu mạnh ở các trường đại học.

Tài liệu tham khảo

- Bakker, M., & Wicherts, J. M. (2014). Outlier removal, sum scores, and the inflation of the type I error rate in independent samples t tests: The power of alternatives and recommendations. *Psychological Methods*, 19(3), 409-427. <https://doi.org/10.1037/met0000014>
- Gigerenzer, G. (2018). Statistical Rituals: The Replication Delusion and How We Got There. *Advances in Methods and Practices in Psychological Science*, 1(2), 198-218. <https://doi.org/10.1177/2515245918771329>
- Hoàng Văn Hào, Phạm Hoàng Điệp (2020). Năng lực nghiên cứu khoa học của giảng viên đại học: Một nghiên cứu đối với nhóm giảng viên giảng dạy các ngành ngoài sư phạm tại Trường Đại học Thủ đô Hà Nội. *Tạp chí Kinh tế và Phát triển*, 273, 93-102.
- Kruppa, J., Rohmann, J., Herrmann, C., Sieg, M., Rubarth, K., & Piper, S. (2021). What statistics instructors need to know about concept acquisition to make statistics stick. *Journal of University Teaching & Learning Practice*, 18(2), 2. <https://ro.uow.edu.au/jutlp/vol18/iss2/02>
- Lai, Z., Xu, Y., Jin, Z., & Zhang, D. (2014). Human gait recognition via sparse discriminant projection learning. In *IEEE Transactions on Circuits and Systems for Video Technology*, 24(10), 1651-1662.
- Levy, P. S., & Lemeshow, S. (2013). *Sampling of populations: Methods and applications*. John Wiley & Sons.
- Nguyễn Thị Mỹ Lộc, Trần Trung, Nguyễn Danh Nam (2023). *Công bố khoa học giáo dục từ phân tích dữ liệu*. NXB Đại học Quốc gia Hà Nội.
- Nguyễn Tiến Trung, Đinh Đức Tài, Phạm Hùng Hiệp, Lương Đình Hải, Phạm Thị Oanh (2024). Nghiên cứu khoa học giáo dục Việt Nam trên cơ sở dữ liệu Scopus trong 10 năm thực hiện Nghị quyết 29-NQ/TW: Xu hướng và hợp tác quốc tế. *Tạp chí Giáo dục*, 24(1), 1-6.
- Vu, Q. H., Tran, T. B. N., Nguyen, L. V. A., Nguyen, T. T. Q., Lai, V. K. T., Nguyen, T. T., & Pham, H. H. (2023). Teaching Mathematics for Business Administration Students in Vietnam: A Curriculum Analysis. *Journal of International Business Education*, 18, 153-170. <https://www.neilsonjournals.com/JIBE/abstractjibe18hiepetal.html>