

SỬ DỤNG MỘT SỐ KIỂM ĐỊNH TRONG PHÂN TÍCH DỮ LIỆU NGHIÊN CỨU KHOA HỌC GIÁO DỤC

Nguyễn Chí Bảo¹,
Trần Trung^{2,+}

¹Trường Đại học Vinh; ²Học viện Dân tộc
+Tác giả liên hệ • Email: trungt1978@gmail.com

Article history

Received: 20/3/2025

Accepted: 14/4/2025

Published: 20/5/2025

Keywords

t-test, McNemar test, Chi-square test, educational data analysis, inferential statistics

ABSTRACT

Inferential statistical tests play a vital role in analyzing educational data by evaluating differences, relationships between variables, and the effectiveness of teaching methods. This research focuses on applying four widely used tests: the independent t-test, paired t-test, McNemar test, and Chi-square test. Using real educational data examples, the study demonstrates how these tests provide scientific evidence to identify differences in mathematics scores between male and female students (independent t-test), assess the effectiveness of differentiated teaching approaches (paired t-test), analyze pass/fail rates after exam preparation (McNemar test), and examine the relationship between academic tracks and university major selections (Chi-square test). The findings underscore the importance of selecting appropriate statistical tests based on research design and transforming analytical results into practical educational interventions. The study highlights how proper application of these inferential tests can enhance educational research quality and inform teaching improvement strategies.

1. Mở đầu

Kiểm định là một phương pháp thống kê suy diễn/suy luận liên quan đến phân tích dữ liệu từ các mẫu, và kết quả thu được được sử dụng để khái quát hóa cho toàn bộ quần thể. Phương pháp này cố gắng đưa ra nhiều suy luận khác nhau về một tập dữ liệu từ một mẫu nhất định. Nhiệm vụ của kiểm định thống kê suy luận là ước lượng, kiểm định giả thuyết và đưa ra quyết định (Neideen & Brasel, 2007). Các nghiên cứu quốc tế như Rahayu và cộng sự (2024) đã chứng minh hiệu quả của phương pháp này trong việc đánh giá tác động của các yếu tố giáo dục. Tuy nhiên, việc nghiên cứu các kiểm định trên còn khá hạn chế và chưa được phổ biến tại Việt Nam, đặc biệt là về giáo dục. Việc lựa chọn mô hình phù hợp không chỉ phụ thuộc vào bản chất của dữ liệu mà còn liên quan đến mục tiêu nghiên cứu và giả định thống kê của từng phương pháp. Trong bài báo này, chúng tôi trình bày một số kiểm định thống kê thường được sử dụng trong phân tích dữ liệu nghiên cứu khoa học giáo dục, cùng với các ứng dụng thực tiễn của chúng. Thông qua việc giới thiệu các kiểm định thống kê, bài báo nhằm cung cấp cho các nhà nghiên cứu giáo dục một cái nhìn tổng quan và công cụ phân tích phù hợp để khai thác dữ liệu một cách hiệu quả. Bài báo này sử dụng các phần mềm thống kê (SPSS, R, JASP, STARA) để phân tích, bạn đọc có thể tham khảo sử dụng ở tài liệu của Field (2024) và Hoàng Trọng và Chu Nguyễn Mộng Ngọc (2005). Nghiên cứu áp dụng phương pháp định lượng thông qua phân tích dữ liệu thống kê từ các kiểm định suy luận (t-test, McNemar, Chi-square), kết hợp với dữ liệu thực nghiệm thu nhập từ các trường THPT/THCS trên địa bàn được trình bày dưới dạng bảng biểu.

2. Kết quả nghiên cứu

2.1. Kiểm định t-test độc lập

Kiểm định t-test thường được sử dụng để so sánh trung bình của hai nhóm. Các nhóm được so sánh có thể độc lập với nhau, chẳng hạn như nam và nữ (Kim, 2019). Trong mục này, chúng tôi sẽ chỉ tập trung vào kiểm định t-test độc lập. Kiểm định t-test độc lập là phương pháp thống kê dùng để so sánh giá trị trung bình của một biến liên tục giữa hai nhóm độc lập. Phương pháp này thường được áp dụng trong các nghiên cứu để xác định sự khác biệt có ý nghĩa thống kê giữa hai nhóm độc lập. Ví dụ, khi nhà nghiên cứu giáo dục hoặc GV muốn so sánh sự khác biệt về điểm môn Toán giữa nam và nữ (biến phụ thuộc là “điểm môn Toán”, biến độc lập “giới tính” gồm hai nhóm: nam và nữ), hay khi so sánh sự khác biệt về sự căng thẳng khi đại học giữa ở thành thị và nông thôn (biến phụ thuộc là “mức độ lo lắng khi thi”, biến độc lập là “vị trí vùng miền” với hai nhóm: thành thị và nông thôn).

Trong kiểm định t-test, khi phương sai của tổng thể bằng nhau nên chỉ cần một ước lượng chung về phương sai cho cả hai nhóm. Phương sai chung này được tính bằng công thức: $s_p^2 = \frac{(n_1 - 1)}{(n_1 - 1) + (n_2 - 1)} s_1^2 + \frac{(n_2 - 1)}{(n_1 - 1) + (n_2 - 1)} s_2^2$, trong đó (s_1^2) và (s_2^2) là phương sai mẫu của hai nhóm. Giá trị thống kê t thu được thay thế cả hai phương sai tổng thể (σ_1^2) và (σ_2^2) bằng phương sai chung (s_p^2) . Bậc tự do (df) của kiểm định này được tính theo công thức $(n_1 + n_2 - 2)$. Công thức kiểm định t-test khi đó được biểu diễn như sau: $t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} = \frac{\bar{X}_1 - \bar{X}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{(n_1 + n_2 - 2)}$.

Trong thực tế, có nhiều trường hợp không thể giả định rằng phương sai bằng nhau giữa hai nhóm. Khi đó, chúng ta phải ước tính chúng riêng biệt và công thức tính thống kê kiểm định được biểu diễn như sau: $t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t_v$,

trong đó v là bậc tự do Satterthwaite tính theo công thức: $v = \frac{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}{\frac{(s_1^2 / n_1)^2}{n_1 - 1} + \frac{(s_2^2 / n_2)^2}{n_2 - 1}}$. Chi tiết về các công thức có

thể được tìm hiểu thêm trong nghiên cứu của Kim (2015).

Ví dụ, khi tiến hành thực nghiệm nghiên cứu liệu có sự khác biệt về điểm môn Toán giữa HS nam và nữ trong một trường học hay không, chúng tôi thu thập điểm thi cuối học kì 1 năm học 2024-2025 theo quy trình chuẩn của nhà trường. Mẫu nghiên cứu gồm 48 HS lớp 10 (24 nam và 24 nữ) được chọn ngẫu nhiên từ danh sách toàn khối bằng phương pháp ngẫu nhiên phân tầng đảm bảo cân bằng giới tính ở một trường THPT trên địa bàn như bên:

Bảng 1. Thống kê điểm môn Toán cuối học kì 1 ở trường THPT

HS nam			HS nữ		
9.25	9	9.5	8.75	8.25	6
8.25	9	8	8.5	7	6
8.5	9.75	9.5	8	7	5.5
7	8.75	10	6	6	5
9	9.5	8.25	8	7	7.5
8	8.5	7.5	9	7	5
8	9	8.25	6	5.5	6.5
7	9.75	9.25	9	7.25	7

Từ dữ liệu trên có thể sử dụng phần mềm thống kê (SPSS, R, JASP, STARA) để phân tích, thu được kết quả cho thấy rằng, thống kê F và giá trị ý nghĩa p -value là 0,076 lớn hơn 0,05 nên phương sai nhóm có thể được coi là bằng nhau. Do phương sai bằng nhau, nên chúng ta sử dụng kết quả t-test ở hàng Equal variances assumed. Giá trị “Sig. (2-tailed)” nhỏ hơn 0,01, nghĩa là các trung bình của nhóm khác nhau có ý nghĩa thống kê. Nhìn vào bảng kết quả của phần mềm thống kê, giá trị Mean Difference là 1.73958 mang ý nghĩa điểm của nam cao hơn so với những nữ là 1,73958 điểm, với khoảng tin cậy 95% từ 1,12757 đến 2,35159. Kết quả nghiên cứu chỉ ra sự khác biệt đáng kể về điểm Toán giữa nam và nữ, nhưng không phải do năng lực bẩm sinh. Tuy nhiên, do giới hạn của thiết kế nghiên cứu mô tả hiện tại, chúng tôi chưa xác định nguyên nhân của sự khác biệt này.

2.2. Kiểm định t-test bất cặp

Kiểm định t-test bất cặp là một công cụ thống kê dùng để ước tính liệu sự khác biệt trung bình giữa hai tập dữ liệu quan sát có bằng 0 hay không. Trong phép kiểm định này, mỗi đối tượng hoặc thực thể được đo lường hai lần, tạo ra hai tập quan sát (Hasijsa, 2023). Giả sử một GV muốn đánh giá hiệu quả của phương pháp giảng dạy mới. Họ có thể tiến hành kiểm tra kiến thức của trước khi áp dụng phương pháp này (điểm kiểm tra đầu kì) và sau khi kết thúc chương trình (điểm kiểm tra cuối kì). Bằng cách sử dụng t-test bất cặp, GV có thể xác định liệu sự cải thiện điểm số trung bình của sau khóa học có ý nghĩa thống kê hay không, hay chỉ là do ngẫu nhiên. Một vài ví dụ khác chẳng hạn như so sánh điểm thi môn Toán của yếu trước và sau khi tham gia phụ đạo, hay có thể khảo sát mức độ tự tin của sinh viên khi thuyết trình trước và sau khóa rèn luyện kỹ năng, đo chỉ số tập trung của tiểu học trước và sau khi áp dụng 15 phút vận động giữa các tiết học, hoặc thậm chí đánh giá hiệu quả của khóa đào tạo GV. Tuy nhiên, khi kiểm định t-test, các ví dụ đều yêu cầu cùng một nhóm đối tượng được đo lường ở hai thời điểm, dữ liệu phải đảm bảo tính phụ thuộc và được phân phối gần như chuẩn, các quan sát trong một điều kiện can thiệp là độc lập với

nhau. Trong kiểm định t-test bất cặp, hiệu số trung bình được tính theo công thức: $\bar{d} = \frac{\sum d}{n}$, trong đó $d = x_1 - x_2$

là hiệu số giữa các cặp quan sát. Khi đó, độ lệch chuẩn được tính: $s_d = \sqrt{\frac{\sum (d - \bar{d})^2}{n-1}}$ và giá trị thống kê $t = \frac{\bar{d}}{s_d} \sqrt{n}$.

Ví dụ thực nghiệm khi đánh giá phương pháp dạy học phân hóa trong môn Toán, nghiên cứu này sử dụng phương pháp thực nghiệm: tiến hành thu thập kết quả với 30 HS lớp 11 trường THPT trên địa bàn trong 6 tuần sau khi triển khai dạy học. Chúng tôi đánh giá kiến thức ban đầu thông qua bài kiểm tra Toán và đánh giá kết quả sau can thiệp bằng bài kiểm tra có cấu trúc tương đương, thu được bảng sau:

Bảng 2. Thống kê điểm môn Toán trước và sau khi áp dụng phương pháp dạy học phân hóa

STT	Điểm trước	Điểm sau	STT	Điểm trước	Điểm sau
1	6.00	8.50	16	6.25	8.00
2	7.75	9.25	17	7.25	9.00
3	7.50	8.50	18	7.00	9.00
4	8.00	9.00	19	7.50	8.50
5	5.00	7.00	20	8.25	8.00
6	9.75	9.50	21	8.75	9.50
7	6.50	7.75	22	8.00	9.50
8	6.25	8.00	23	7.00	8.00
9	7.75	9.00	24	7.25	8.50
10	7.00	8.00	25	6.00	8.00
11	7.25	9.00	26	4.50	7.50
12	9.50	10.00	27	6.50	8.00
13	9.25	9.50	28	7.50	8.25
14	7.25	9.00	29	4.00	7.00
15	7.00	8.25	30	8.00	9.25

Giả thuyết không (H_0): Không có sự khác biệt về điểm trung bình trước và sau. Giả thuyết đối (H_1): Có sự khác biệt. Từ dữ liệu trên có thể sử dụng phần mềm thống kê (SPSS, R, JASP, STARA) để phân tích, thu được kết quả cho thấy giá trị “Sig.(2-tailed)” nhỏ hơn 0,01, nghĩa là các trung bình của nhóm khác nhau có ý nghĩa thống kê. Ngoài ra, giá trị “Mean = 1,35833” chứng tỏ khi sử dụng phương pháp dạy học môn Toán, điểm số sau cao hơn điểm số trước là 1,35833 điểm, với khoảng tin cậy từ 1,06963 đến 1,64704, $t(29df) = 9,623$. Điểm trung bình trước là 7,1833 và sau là 8,5417. Như vậy, phương pháp dạy học mới đã cải thiện điểm số toán đáng kể ($p < 0,001$), đặc biệt hiệu quả đối với HS có học lực trung bình (điểm trước < 7). Tuy nhiên, vẫn có một vài em điểm thấp hơn trước (2/30 HS), có thể do chưa thích ứng phương pháp mới hoặc yếu tố về độ khó của bài kiểm tra,... Tổng kết, phương pháp này có hiệu quả, nhưng cần linh hoạt điều chỉnh theo điều kiện người học và bối cảnh cụ thể nhằm phát huy hiệu quả tối đa hơn nữa:

2.3. Kiểm định McNemar

Kiểm định được mô tả chính xác nhất là một bảng phân loại chéo 2×2 của các cặp phản hồi được ghép đôi (hoặc so khớp) đối với một biến nhị phân. Nói một cách đơn giản, kiểm định McNemar có thể được xem như một dạng kiểm định chi-square sử dụng dữ liệu phụ thuộc (tức có tương quan hoặc ghép đôi) thay vì các mẫu độc lập (không liên quan) (Adedokun & Burgess, 2012). Kiểm định McNemar là một kiểm định thống kê phi tham số; nghĩa là nó không phụ thuộc vào phân phối và có thể áp dụng cho các tập dữ liệu hoặc mẫu không tuân theo phân phối chuẩn (Ciechalski et al., 2022).

Ví dụ minh họa: Ứng dụng kiểm định McNemar để đánh giá hiệu quả ôn thi môn Toán vào lớp 10 của một trường THCS trên địa bàn khi so sánh kết quả giữa lần 1 là thi thử và lần 2 là thi chính thức. Dữ liệu được lấy từ toàn bộ 260 em HS khối 9 của trường THCS đó qua kết quả của bài thi thử và thi chính thức năm học 2023-2024. Các kết quả được mã hóa thành giá trị 0 là Không đạt (< 5 điểm), và 1 là Đạt (≥ 5 điểm). Khi đó, giả thuyết không (H_0): Tỷ lệ đạt không thay đổi sau ôn tập và (H_1): Tỷ lệ đạt thay đổi. Mức ý nghĩa alpha được chọn là 5%. Dữ liệu được trình bày ở bảng sau:

Bảng 3. Thống kê điểm môn Toán thi thử và thi chính thức

Thi thử	Thi chính thức		
		Đạt (1)	Không đạt (0)
	Đạt (1)	a = 80	b = 30
	Không Đạt (0)	c = 90	d = 60
		$n_2 = 170$	

Trong đó, a là số HS đạt cả 2 lần; b là số HS đạt ở thi thử nhưng không đạt ở thi chính thức; c là số HS đạt ở thi chính thức nhưng lại không đạt ở thi thử; d là số HS không đạt ở 2 lần. Từ dữ liệu trên có thể sử dụng phần mềm thống kê (SPSS, R, JASP, STARA) để phân tích, thu được kết quả cho thấy, trung vị điểm thi chính là 1, cao hơn trung vị đề thi thử là 1 điểm. Giá trị chính xác của $p < 0,001$ cho thấy sự khác biệt giữa thi thử và thi chính thức. Hay nói khác đi, chương trình ôn tập có hiệu quả rõ rệt. Tỷ lệ đạt tăng từ 42% lên 65%. Ngoài ra, có 34,6% HS cải thiện từ không đạt lên đạt. Điều đó chứng tỏ nên tiếp tục duy trì phương pháp ôn tập hiện tại, tuy nhiên cần phải chú ý, tập trung hỗ trợ 30 HS bị giảm điểm. Đồng thời, cần phải phân tích sâu nguyên nhân cải thiện và giảm sút để có kết quả tốt nhất.

2.4. Kiểm định Chi-square

Kiểm định Chi-square (χ^2) là một phương pháp thống kê phi tham số được sử dụng rộng rãi để đánh giá mối quan hệ giữa các biến phân loại (nominal hoặc categorical). Đặc trưng của dữ liệu định danh là các giá trị thuộc về các nhóm riêng biệt và không chồng lấn lên nhau (Ugoni & Walker, 1995). Phương pháp này thường được áp dụng khi dữ liệu được trình bày dưới dạng bảng tần số (frequency table). Có hai dạng kiểm định Chi-square phổ biến:

- Kiểm định tính độc lập: Dùng để xác định xem có mối liên hệ giữa hai biến định tính hay không. Ví dụ: Kiểm tra xem giới tính (Nam/Nữ) có liên quan đến việc lựa chọn ngành học (Khoa học/Xã hội) hay không. Đặc điểm của kiểm định này là tổng các hàng và cột trong bảng tần số là ngẫu nhiên.

- Kiểm định tính đồng nhất: Dùng để so sánh sự phân bố tỉ lệ giữa các nhóm khác nhau, chẳng hạn so sánh tỉ lệ HS đạt loại giỏi giữa hai trường THPT khác nhau. Đặc điểm của kiểm định này là tổng của một biến (hàng hoặc cột) được cố định trước.

Mặc dù cách tính toán và diễn giải kết quả của hai kiểm định này tương tự nhau, nhưng chúng khác biệt về mục đích sử dụng và thiết kế nghiên cứu. Độ chính xác của kiểm định phụ thuộc vào kích thước mẫu và phân phối tần số trong từng ô của bảng. Đặc biệt, với bảng 2x2 (2 hàng và 2 cột), cần đảm bảo rằng tần số kì vọng trong mỗi ô không quá nhỏ (thường ≥ 5) để kết quả đáng tin cậy. Chi-square cung cấp giá trị p-value để xác định tính ý nghĩa thống kê. Nếu p-value < 0.05 , ta có thể kết luận có sự khác biệt hoặc mối liên hệ giữa các biến. Tuy nhiên, cần lưu ý rằng kiểm định này không chỉ ra mối quan hệ nhân quả mà chỉ phản ánh sự tương quan. Giá trị thống kê χ^2 được xác định bằng cách so sánh giữa tần số quan sát (O - Observed frequency) và tần số kì vọng (E - Expected frequency) theo

công thức sau $\chi^2 = \sum \frac{(O - E)^2}{E}$ (Ugoni & Walker, 1995).

Thực nghiệm ở trường phổ thông trên địa bàn, nghiên cứu sử dụng phương pháp điều tra cắt ngang để hiểu rõ xu hướng chọn ngành thi đại học của HS theo ban (Tự nhiên/Xã hội) để có định hướng tư vấn, thiết kế chương trình học tập hiệu quả, toàn bộ 605 em HS khối 12 được chọn và khảo sát nguyện vọng ngành thi đại học từ danh sách cố sẵn. Dữ liệu được thu thập thông qua hai nguồn chính: (1) Phiếu đăng kí nguyện vọng thi đại học chính thức của nhà trường, trong đó HS tự khai báo ngành học dự định; (2) Hồ sơ học tập để xác định ban học của từng HS. Công cụ thu thập là bảng hỏi tiêu chuẩn được thiết kế gồm 7 nhóm ngành học chính. Quy trình thực hiện bao gồm: rà soát hồ sơ đăng kí nguyện vọng, đối chiếu với danh sách lớp để xác định ban học, và nhập liệu vào hệ thống theo các nhóm ngành đã phân loại. Toàn bộ quá trình thu thập dữ liệu được thực hiện trong vòng 2 tuần trước kì thi đại học để đảm bảo tính cập nhật của thông tin. Kết quả được minh họa ở bảng sau:

Bảng 4. Thống kê ngành học và ban học của HS lớp 12

Ban học Ngành học	Khoa học tự nhiên	Kĩ thuật	Kinh tế	Xã hội nhân văn	Y dược	Nghệ thuật	Khác	Tổng
Tự nhiên	80	110	69	35	25	15	9	343
Xã hội	19	15	86	78	29	25	10	262
Tổng	99	125	155	113	54	40	19	605

Giả thuyết không (H_0): Không có mối liên hệ giữa ban học (Tự nhiên/Xã hội) và ngành thi đại học, giả thuyết đối (H_1): có mối liên hệ giữa ban học và ngành thi đại học. Từ dữ liệu trên có thể sử dụng phần mềm thống kê (SPSS, R, JASP, STARA) để phân tích, thu được kết quả Chi-square tests, giá trị $\chi^2 = 122.208$ với bậc tự do bằng 6, p-value bé hơn 0,05 nên bác bỏ H_0 , tức là có bằng chứng thống kê để khẳng định mối liên hệ giữa ban học và ngành thi đại học với độ tin cậy 95%. Hệ số Cramer's V = 0,449 thể hiện mối liên hệ mạnh đó. Tất cả các ô trong bảng có tần số kì vọng lớn hơn 5 nên dữ liệu phù hợp để sử dụng kiểm định Chi-square. Tính phần trăm chênh lệch giữa quan sát và kì vọng, ta có bảng sau giúp nhà trường định hướng rõ hơn cho chương trình giảng dạy của mình.

3. Kết luận

Trong nghiên cứu khoa học giáo dục, việc lựa chọn kiểm định thống kê cần căn cứ vào ba yếu tố then chốt: thiết kế nghiên cứu, loại biến số và mục tiêu phân tích. Chẳng hạn, nếu nghiên cứu nhằm so sánh sự khác biệt giữa hai nhóm độc lập, như HS nam và nữ, thì t-test độc lập là lựa chọn thích hợp, trong khi t-test bất cặp được dùng khi đánh giá hiệu quả can thiệp trên cùng một nhóm. Đối với dữ liệu nhị phân (đạt/không đạt) thu thập từ cùng đối tượng ở hai thời điểm, phân tích các cặp phù hợp, McNemar là lựa chọn phù hợp hơn, còn Chi-square thích hợp để phân tích mối liên hệ giữa các biến phân loại. Mỗi kiểm định đều có những ưu và nhược điểm riêng. Các kiểm định t-test (yêu cầu phân phối gần chuẩn và phương sai đồng nhất) cung cấp thông tin định lượng chính xác nhưng kém linh hoạt với dữ liệu không chuẩn. Ngược lại, McNemar và Chi-square (không yêu cầu phân phối chuẩn) phù hợp với dữ liệu định tính hoặc phân loại nhưng cần lưu ý: (1) McNemar không áp dụng cho biến nhiều mức (nominal nhiều hơn 2 nhóm) trừ khi được gộp về 2 nhóm, (2) Chi-square đòi hỏi tần số ô ≥ 5 để đảm bảo độ tin cậy của kết quả. Một nghiên cứu có thể kết hợp nhiều kiểm định nhưng không nhất thiết phải sử dụng tất cả, ví dụ: dùng t-test để so sánh điểm số, sau đó áp dụng Chi-square phân tích mối tương quan giữa kết quả học tập và các yếu tố khác. Do đó, để đưa ra những kết luận chính xác và có giá trị thực tiễn trong giáo dục, nhà nghiên cứu cần hiểu rõ bản chất dữ liệu, mục tiêu nghiên cứu cũng như điều kiện áp dụng của từng kiểm định, từ đó lựa chọn công cụ phân tích thống kê phù hợp nhất. Tuy nhiên, trong thực tế, còn cần quan tâm đến các yếu tố ngoại biến như điều kiện gia đình, động lực học tập, tác động đồng thời của cấp độ cá nhân, lớp học, nhà trường,... Trong các nghiên cứu tiếp theo, có thể mở rộng hướng tiếp cận bằng cách kết hợp mô hình hồi quy đa biến và phương pháp hỗn hợp, từ đó phân tích sâu hơn ảnh hưởng tổng hợp của cá nhân, gia đình và môi trường giáo dục,... nhằm đề xuất các giải pháp can thiệp toàn diện và hiệu quả hơn.

Tài liệu tham khảo

- Adedokun, O. A., & Burgess, W. D. (2012). Analysis of paired dichotomous data: A gentle introduction to the McNemar test in SPSS. *Journal of MultiDisciplinary Evaluation*, 8(17), 125-131.
- Ciechalski, J. C., Pinkney, J. W., & Weaver, F. S. (2002). *A Method for Assessing Change in Attitude: The McNemar Test*. Poster presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.
- Field, A. (2024). *Discovering statistics using IBM SPSS statistics*. Sage Publications Limited.
- Hasija, Y. (2023). Chapter 3 - Statistical methods in bioinformatics. In: *All About Bioinformatics* (pp. 43-75). Academic Press.
- Hoàng Trọng, Chu Nguyễn Mộng Ngọc (2005). *Phân tích dữ liệu nghiên cứu với SPSS*. NXB Thống kê.
- Kim, T. K. (2015). T test as a parametric statistic. *Korean Journal of Anesthesiology*, 68(6), 540-546.
- Kim, H. Y. (2019). Statistical notes for clinical researchers: the independent samples t-test. *Restorative Dentistry & Endodontics*, 44(3), e26.
- Neideen, T., & Brasel, K. (2007). Understanding statistical tests. *Journal of Surgical Education*, 64(2), 93-96.
- Rahayu, N. I., Nandiyanto, A. B. D., Muktiarni, M., & Hidayat, Y (2024). An Application of Statistical Testing: A Guide to Basic Parametric Statistics in Educational Research Using SPSS. *ASEAN Journal of Science and Engineering*, 4(3), 569-582.
- Ugoni, A., & Walker, B. F. (1995). The Chi square test: an introduction. *COMSIG Review*, 4(3), 61.