

**ORGANIZING REAL-WORLD DATA
EXPLORATION ACTIVITIES
FOR HIGH SCHOOL STUDENTS
WITH THE SUPPORT OF STATISTICAL
EDUCATION SOFTWARE**

Pham The Quan

*Faculty of Mathematics, Hanoi Pedagogical
University 2, Vinh Phuc province, Vietnam*

e-mail: phamthequan@hpu2.edu.vn

Received January 5, 2025.

Revised January 20, 2025.

Accepted January 27, 2025.

**TỔ CHỨC HOẠT ĐỘNG KHÁM PHÁ
DỮ LIỆU THỰC TẾ CHO HỌC SINH
Ở TRƯỜNG TRUNG HỌC PHỔ THÔNG
VỚI SỰ HỖ TRỢ CỦA PHẦN MỀM
GIÁO DỤC THỐNG KÊ**

Phạm Thế Quân

*Khoa Toán, Trường Đại học Sư phạm Hà Nội 2,
tỉnh Vĩnh Phúc, Việt Nam*

e-mail: phamthequan@hpu2.edu.vn

Ngày nhận bài: 5/1/2025.

Ngày sửa bài: 20/1/2025.

Ngày nhận đăng: 27/1/2025.

Abstract. The organization of real-world data exploration activities with the support of statistical education software helps students gain a deeper understanding of statistical concepts and develop statistical reasoning skills. The article introduces several statistical education software tools including CODAP, a support tool for organizing real-world data exploration activities for high school students. The proposed process for organizing these activities consists of two phases: (1) Preparation, including familiarization with the software, data selection, and question development; (2) Implementation of data exploration activities, involving the steps of problem identification, solution exploration, data analysis, conclusion, and application. Specific illustrative examples are provided to support teachers in effectively applying these activities in real-world teaching contexts.

Keywords: teaching statistics, real-world contexts, real-world data, statistical education software, statistical reasoning.

Tóm tắt. Việc tổ chức hoạt động khám phá dữ liệu thực tế với sự hỗ trợ của phần mềm giáo dục thống kê giúp học sinh hiểu sâu hơn các khái niệm thống kê và phát triển khả năng suy luận thống kê. Bài báo giới thiệu một số phần mềm giáo dục thống kê, trong đó CODAP được sử dụng như một công cụ hỗ trợ tổ chức các hoạt động khám phá dữ liệu thực tế cho học sinh ở trường trung học phổ thông. Quy trình tổ chức hoạt động khám phá dữ liệu thực tế được đề xuất gồm hai giai đoạn: (1) Chuẩn bị, bao gồm làm quen với phần mềm, lựa chọn dữ liệu và xây dựng câu hỏi; (2) Triển khai hoạt động khám phá dữ liệu, với các bước: tìm hiểu vấn đề, tìm giải pháp, phân tích dữ liệu, kết luận và vận dụng. Bài báo cũng đưa ra các ví dụ minh họa cụ thể nhằm hỗ trợ giáo viên áp dụng hiệu quả trong thực tế giảng dạy.

Từ khóa: dạy học thống kê, bối cảnh thực, dữ liệu thực tế, phần mềm giáo dục thống kê, suy luận thống kê.

1. Mở đầu

Trong bối cảnh cách mạng công nghiệp 4.0 và quá trình chuyển đổi số diễn ra mạnh mẽ, dữ liệu ngày càng được coi là một nguồn tài nguyên quan trọng, đóng vai trò nền tảng trong nhiều lĩnh vực. Việc trang bị cho học sinh (HS) kiến thức về thống kê, xác suất và khả năng làm việc với dữ liệu đã

trở thành một yêu cầu cấp thiết trong giáo dục hiện đại. Phân tích dữ liệu không chỉ là công cụ hỗ trợ đưa ra các quyết định, mà còn là phương tiện giúp HS hiểu sâu hơn về các vấn đề đa dạng của xã hội, kinh tế và khoa học. Bản chất của Thống kê đòi hỏi sự gắn kết chặt chẽ với bối cảnh. Như Cobb và Moore (1997) đã nhấn mạnh, “Dữ liệu không chỉ là con số, mà là con số trong bối cảnh” [1]. Điều này ngụ ý rằng để hiểu và phân tích dữ liệu một cách chính xác, HS cần nắm rõ bối cảnh xuất phát của dữ liệu. Không giống như trong Toán học, nơi các khái niệm và quy tắc có thể được áp dụng một cách phổ quát mà không cần quan tâm đến bối cảnh cụ thể, việc áp dụng các phương pháp thống kê đòi hỏi sự hiểu biết sâu sắc về tình huống thực tế.

Ở Việt Nam, việc tích hợp công nghệ vào dạy học nói chung và dạy học toán nói riêng đang nhận được sự quan tâm lớn [2]. Chương trình Giáo dục phổ thông 2018 đã đưa ra các yêu cầu về việc tăng cường tính ứng dụng thực tế và vận dụng công nghệ trong các môn học, trong đó có môn Toán [3]. Các nghiên cứu đã chỉ ra rằng việc sử dụng công cụ công nghệ trong giảng dạy thống kê và xác suất có thể cải thiện đáng kể suy luận thống kê và tư duy thống kê của HS [4], [5]. Theo Pratt và cộng sự (2011), công nghệ giúp HS tiếp cận với các bộ dữ liệu thực tế một cách trực quan, hỗ trợ việc phân tích và ra quyết định dựa trên dữ liệu [6]. Đặc biệt, các phần mềm giáo dục thống kê như Fathom, TinkerPlots, CODAP không chỉ cung cấp các công cụ mạnh mẽ để khám phá dữ liệu mà còn khuyến khích HS tự chủ trong quá trình học tập, từ đó phát triển suy luận thống kê và tư duy thống kê. Tuy nhiên, việc áp dụng công nghệ trong dạy học thống kê và xác suất vẫn còn gặp nhiều hạn chế, từ khả năng tiếp cận công cụ của giáo viên (GV) đến việc tổ chức các hoạt động học tập phù hợp với HS.

Bài báo này nhằm giải quyết các thách thức trên bằng cách đề xuất quy trình tổ chức hoạt động khám phá dữ liệu thực tế với phần mềm giáo dục thống kê, từ đó góp phần nâng cao hiệu quả dạy học thống kê trong các trường Trung học phổ thông. Bài báo cung cấp một cách tiếp cận mới để giảng dạy nội dung này, giúp HS không chỉ nắm bắt kiến thức mà còn phát triển khả năng làm việc với dữ liệu và giải quyết các vấn đề thực tiễn.

2. Nội dung nghiên cứu

2.1. Vai trò của bối cảnh trong dạy học thống kê

Mục tiêu cốt lõi của thống kê là hỗ trợ con người hiểu sâu hơn về các hiện tượng trong thế giới thực [7]. Quá trình khai thác thông tin từ dữ liệu nhằm làm sáng tỏ các vấn đề thực tiễn giữ vai trò quan trọng trong việc đưa ra các quyết định dựa trên bằng chứng. Phân tích dữ liệu không chỉ giới hạn ở việc áp dụng các công thức toán học mà còn yêu cầu sự am hiểu sâu sắc về bối cảnh mà dữ liệu được thu thập. Hiểu biết về bối cảnh không chỉ giúp xác định các câu hỏi nghiên cứu một cách phù hợp mà còn đảm bảo lựa chọn phương pháp phân tích chính xác và diễn giải kết quả một cách hợp lý. Theo Wild & Pfannkuch (1999), ba yếu tố nền tảng trong phân tích dữ liệu: kiến thức thống kê, hiểu biết về bối cảnh và dữ liệu - là những thành phần không thể tách rời, cùng đóng góp vào việc tạo ra các kết luận đáng tin cậy và có giá trị thực tiễn [7].

Trong quá trình giải quyết các vấn đề thống kê, ngữ cảnh dữ liệu đóng vai trò then chốt và không thể thiếu. Ngữ cảnh này đại diện cho bối cảnh thực tế nơi vấn đề phát sinh, đồng thời có mối liên hệ chặt chẽ với việc đưa ra lời giải hoặc hiểu sâu hơn về tình huống đang được nghiên cứu. Ngữ cảnh dữ liệu bao hàm kiến thức chuyên môn liên quan đến lĩnh vực của dữ liệu, cũng như các thông tin về cách thức thu thập và xử lý dữ liệu. Nói cách khác, ngữ cảnh dữ liệu không chỉ giúp xây dựng bức tranh toàn cảnh của vấn đề mà còn định vị dữ liệu trong mối quan hệ với thực tiễn, từ đó làm sáng tỏ ý nghĩa và giá trị của dữ liệu trong việc phân tích và đưa ra quyết định.

Khi phân tích sự khác biệt giữa Toán học và Thống kê, nhiều nhà nghiên cứu đã chỉ ra rằng bối cảnh là một điểm quan trọng cần được nhấn mạnh. Với Toán học thuần túy, bối cảnh thường được sử dụng như điểm khởi đầu để thu hút sự quan tâm của HS trước khi chuyển sang các khái niệm trừu tượng. Các nhà toán học tập trung vào cấu trúc logic, mối quan hệ giữa các đối tượng toán học và tính

nhất quán của hệ thống. Trái lại, Thống kê luôn gắn liền với bối cảnh cụ thể của dữ liệu và vấn đề đang được nghiên cứu [8], [9]. Cobb & Moore (1997) đã nhấn mạnh tầm quan trọng của bối cảnh trong Thống kê: “Dữ liệu không chỉ là con số, mà là con số trong bối cảnh” [1], thống kê tìm kiếm ý nghĩa bằng cách đặt dữ liệu vào bối cảnh. Điều này có nghĩa là việc hiểu và diễn giải dữ liệu thống kê đòi hỏi không chỉ kỹ năng tính toán mà còn cần sử dụng kiến thức về lĩnh vực, hiểu được nguồn gốc của dữ liệu và xem xét các ý nghĩa thực tế của kết quả thống kê.

Việc giảng dạy thống kê trong các bối cảnh thực tiễn có vai trò quan trọng trong việc nâng cao động lực học tập và sự tham gia tích cực của HS. Garfield & Ben-Zvi (2007) cho rằng khi HS được làm việc với dữ liệu thực tế và giải quyết các vấn đề có ý nghĩa, họ thường tham gia vào quá trình học tập một cách tích cực và chủ động hơn [10]. Bối cảnh thực tiễn không chỉ làm cho các bài toán thống kê trở nên sinh động và hấp dẫn hơn mà còn khơi dậy sự hứng thú và tính tò mò của HS. Thông qua việc tham gia vào các vấn đề thực tế, HS có cơ hội kết nối kiến thức thống kê với thế giới xung quanh, từ đó tăng cường sự hiểu biết sâu sắc và phát triển tư duy phân tích.

Vai trò quan trọng của bối cảnh gợi ý rằng việc dạy học thống kê không nên chỉ tập trung vào công thức và các kỹ thuật tính toán, mà cần đặt các bài toán và dữ liệu vào các tình huống thực tế, phù hợp và có ý nghĩa đối với HS. Cách tiếp cận này không chỉ giúp HS phát triển khả năng suy luận thống kê mà còn hỗ trợ họ hiểu rõ hơn ý nghĩa và những hạn chế của các kết quả thống kê khi áp dụng trong các bối cảnh khác nhau. Qua đó, HS có thể vận dụng kiến thức thống kê một cách linh hoạt và hiệu quả hơn trong thực tiễn.

2.2. Sử dụng phần mềm giáo dục thống kê trong dạy học

Trong kỉ nguyên số, công nghệ đã trở thành một yếu tố không thể thiếu trong việc thực hiện các phân tích thống kê [11]. Các phần mềm hiện đại cho phép xử lí dữ liệu lớn, nhanh chóng và chính xác, từ thu thập đến trực quan hóa. Công nghệ mở rộng phạm vi phân tích, giúp khám phá các mối quan hệ và xu hướng phức tạp mà phương pháp truyền thống khó nhận diện. Công nghệ không chỉ cải thiện hiệu quả và độ chính xác trong phân tích thống kê mà còn mở ra nhiều cơ hội sáng tạo trong học tập, nghiên cứu và ứng dụng thực tiễn.

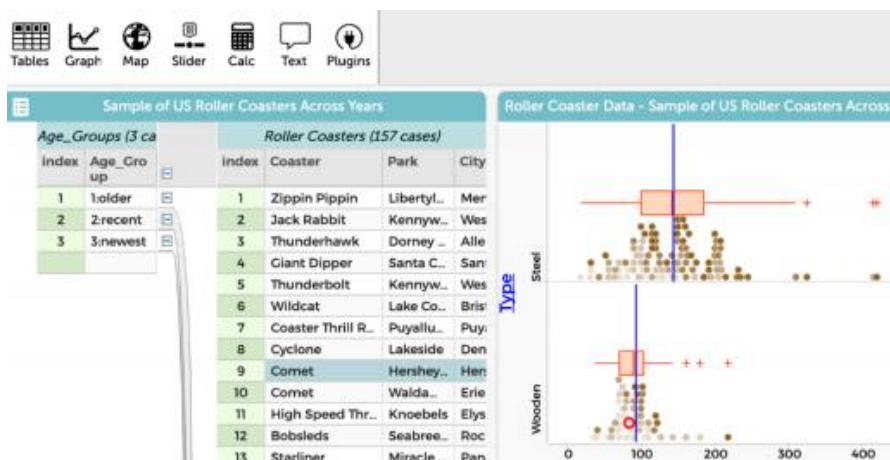
Công nghệ đã mang lại những ảnh hưởng sâu rộng đến phương thức làm việc của các nhà thống kê do đó thay đổi cách dạy và học thống kê. Trong các thập kỉ 1960 và 1970, những cuộc thảo luận sôi nổi đã diễn ra nhằm định hướng việc hiện đại hóa giáo dục thống kê trong bối cảnh công nghệ phát triển mạnh mẽ [12]. Trong đó, hai quan điểm chính được hình thành: Quan điểm thứ nhất cho rằng công nghệ trong học tập thống kê cần phản ánh lí thuyết và thực hành của các công cụ thống kê chuyên nghiệp, nhằm thu hẹp tối đa khoảng cách giữa việc học và ứng dụng các phương pháp thống kê vào thực tiễn. Theo quan điểm này, việc tích hợp các phần mềm chuyên dụng như SAS, SPSS, STATA hay R vào giảng dạy được ủng hộ vì khả năng mô phỏng môi trường làm việc thực tế của các nhà thống kê chuyên nghiệp. Ngược lại, quan điểm thứ hai tập trung vào việc phát triển các công cụ công nghệ được thiết kế riêng cho giáo dục, với mục tiêu tối ưu hóa quá trình học tập Thống kê. Các nhà giáo dục ủng hộ quan điểm này cho rằng, để khai thác hiệu quả công nghệ trong trường học, cần xây dựng các công cụ phù hợp với nhu cầu của HS chưa có kinh nghiệm, đồng thời hỗ trợ sự phát triển năng lực của họ theo từng giai đoạn khi họ dần tích lũy kiến thức chuyên môn. Các công cụ như vậy đã được phát triển như:

- + ProbSim và DataScope (những năm 1990): Mô phỏng xác suất và phân tích dữ liệu đơn giản.
- + TableTop (1995) và InspireData: Biểu diễn dữ liệu trên bảng ảo, giúp HS khám phá dữ liệu.
- + Minitools (những năm 2000): Môi trường thử nghiệm hạn chế, hỗ trợ phát triển ý tưởng qua hiển thị trực quan và so sánh dữ liệu.

Năm 1997, Biehler đã đề xuất phát triển một loại phần mềm mới nhằm giải quyết ba vấn đề chính trong giáo dục thống kê [13]: Thứ nhất là sự phức tạp của công cụ thống kê. Các công cụ thống kê hiện có quá phức tạp và khó sử dụng. Thứ hai là hạn chế của phần mềm mô phỏng. Các phần mềm

mô phỏng đóng không đáp ứng đủ nhu cầu giảng dạy, đặc biệt khi yêu cầu sự tương tác và tính linh hoạt trong việc khám phá dữ liệu. Thứ ba là sự phụ thuộc vào nhiều công cụ. Việc phải sử dụng nhiều công cụ khác nhau dẫn đến sự phức tạp không cần thiết, làm giảm hiệu quả và tính nhất quán trong quá trình dạy học. Đề xuất của Biehler đã được hiện thực hóa thông qua hai phần mềm giáo dục thống kê nổi bật: TinkerPlots và Fathom, được phát triển vào đầu những năm 2000. Cả hai công cụ này đã chứng minh được tính hiệu quả trong việc hỗ trợ giảng dạy và học tập thống kê ở nhiều cấp độ và vẫn đang được sử dụng hiện nay. Dự án phát triển phần mềm Fathom được khởi xướng bởi William Finzer vào năm 2002. Song song đó, Key Curriculum Press dưới sự dẫn dắt của Clifford Konold đã cho ra đời phần mềm TinkerPlots vào năm 2005. Cả hai dự án đều hướng tới mục tiêu thiết kế các công cụ hỗ trợ trực quan và phân tích dữ liệu hiệu quả cho HS. Fathom và TinkerPlots đều sở hữu các tính năng cơ bản tương đồng: khả năng lưu trữ, xử lý, tính toán dữ liệu số; xây dựng các loại biểu đồ đa dạng và mô phỏng trực quan các hiện tượng thống kê. Đặc biệt, cả hai phần mềm đều cung cấp công cụ thanh trượt, giúp HS khám phá sự thay đổi của các biến số và quan sát tác động của chúng lên kết quả.

Năm 2016, các nhà phát triển của hai phần mềm thống kê Fathom và TinkerPlots, đã cùng hợp tác để tạo ra CODAP (Common Online Data Analysis Platform). Đây là một nền tảng phân tích dữ liệu trực tuyến, mã nguồn mở, được thiết kế đặc biệt cho môi trường giáo dục. CODAP kế thừa và phát triển những ưu điểm của các phần mềm tiền nhiệm, đồng thời mang đến nhiều cải tiến đáng kể: giao diện trực quan, dễ sử dụng, tính năng phong phú và hoàn toàn miễn phí. Hiện tại, CODAP vẫn đang được cộng đồng phát triển không ngừng để đáp ứng nhu cầu ngày càng cao của việc dạy và học thống kê.



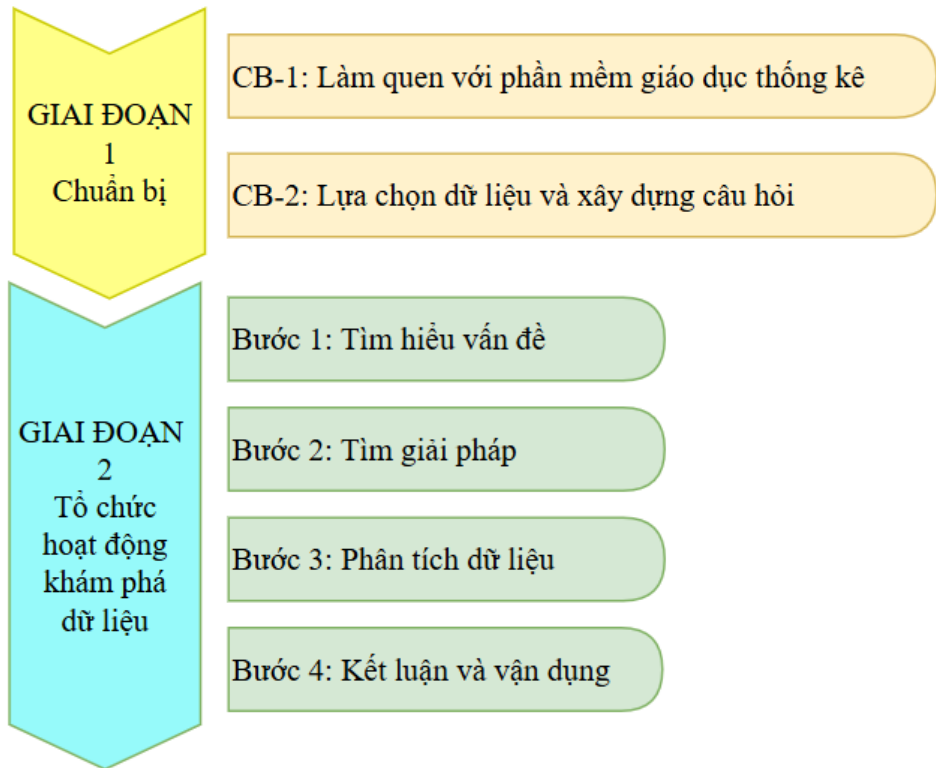
Hình 1. Giao diện của CODAP

Các phần mềm giáo dục thống kê đóng vai trò quan trọng trong việc hỗ trợ quá trình dạy và học thống kê [14], cụ thể:

- + Tự động hóa tính toán: Công nghệ giúp giảm thiểu thời gian thực hiện các phép tính phức tạp, nâng cao độ chính xác và cho phép HS tập trung vào việc hiểu bản chất các khái niệm thống kê.
- + Trực quan hóa khái niệm: Nhờ các hình ảnh động, mô hình tương tác, các khái niệm trừu tượng trong thống kê trở nên dễ hiểu và ghi nhớ hơn.
- + Tăng cường khả năng khám phá dữ liệu: Các phần mềm cung cấp đa dạng các công cụ trực quan hóa dữ liệu, khuyến khích HS chủ động tìm tòi, phân tích và rút ra kết luận từ dữ liệu một cách sâu sắc.
- + Kết nối lý thuyết với thực tiễn: HS có cơ hội làm việc với các tập dữ liệu lớn, thực tế, qua đó rèn luyện kỹ năng thu thập, xử lý và phân tích dữ liệu, đồng thời áp dụng kiến thức đã học vào giải quyết các vấn đề thực tế.

2.3. Quy trình tổ chức hoạt động khám phá dữ liệu thực tế

Việc tổ chức cho HS khám phá dữ liệu thực tế với phần mềm giáo dục thống kê được tiến hành theo hai giai đoạn với trình tự như mô tả ở Hình 2.



Hình 2. Quy trình tổ chức hoạt động khám phá dữ liệu thực tế

* **Giai đoạn 1: Chuẩn bị (CB)**

CB-1: Làm quen với phần mềm giáo dục thống kê

Mục tiêu chính của bước này là giúp HS tiếp cận và làm quen các phần mềm giáo dục thống kê như CODAP hoặc TinkerPlots. GV có thể giới thiệu phần mềm thông qua các bài học trên lớp, giải thích giao diện và chức năng chính của phần mềm, sau đó hướng dẫn HS thực hành những thao tác cơ bản. Ngoài ra, GV có thể giao các nhiệm vụ đơn giản liên quan đến dữ liệu, khuyến khích HS sử dụng phần mềm để áp dụng những kiến thức đã học.

CB-2: Lựa chọn dữ liệu và xây dựng câu hỏi

Đây là bước chuẩn bị quan trọng của GV, nhằm đảm bảo HS có nguồn dữ liệu thực tế để phân tích. GV cần chọn những bộ dữ liệu phù hợp cao với nội dung bài học và trình độ của HS. Một số nguồn mở đáng tin cậy mà GV có thể sử dụng để lựa chọn dữ liệu thực tế được giới thiệu trong Bảng 1. Bên cạnh đó, GV cần thiết kế các câu hỏi định hướng để dẫn dắt HS trong quá trình khám phá dữ liệu. Các câu hỏi có thể tập trung vào việc tìm hiểu đặc điểm của dữ liệu, xác định mối quan hệ giữa các biến số hoặc giải quyết một vấn đề thực tiễn nào đó. Cần chú ý đưa ra câu hỏi thúc đẩy việc sử dụng các kỹ năng suy luận thống kê như mô tả dữ liệu, so sánh, suy luận, xác định mối quan hệ giữa các biến. Có thể đưa ra câu hỏi mở để khuyến khích HS đưa ra nhiều giả thuyết và cách giải thích khác nhau.

Bảng 1. Một số nguồn dữ liệu mở

Stt	Địa chỉ	Thông tin
1	Gapminder (https://www.gapminder.org).	Gapminder cung cấp các bộ dữ liệu phong phú về các chủ đề toàn cầu như dân số, kinh tế, y tế, giáo dục, và môi trường.
2	Our World in Data (https://ourworldindata.org).	Cung cấp dữ liệu và biểu đồ về các vấn đề toàn cầu, bao gồm biến đổi khí hậu, sức khỏe, năng lượng, giáo dục và nghèo đói. Dữ liệu có thể được tải về ở nhiều định dạng khác nhau.
3	Kaggle Datasets (https://www.kaggle.com/datasets).	Kaggle là một nền tảng dành cho khoa học dữ liệu với hàng nghìn bộ dữ liệu thuộc nhiều lĩnh vực.
4	World Bank Open Data (https://data.worldbank.org).	Ngân hàng Thế giới cung cấp dữ liệu kinh tế, xã hội, và môi trường từ hơn 200 quốc gia.
5	Google Public Data Explorer (https://www.google.com/publicdata/directory).	Tổng hợp các bộ dữ liệu từ các nguồn uy tín như Ngân hàng Thế giới, Eurostat, và OECD.
6	Tổng cục thống kê Việt Nam (https://www.gso.gov.vn/).	Các bộ dữ liệu của Việt Nam thuộc nhiều lĩnh vực.

*** Giai đoạn 2: Tổ chức hoạt động khám phá dữ liệu**

Bước 1: Tìm hiểu vấn đề

HS tìm hiểu bối cảnh của vấn đề, thông tin cơ bản về bộ dữ liệu như kích thước, loại dữ liệu, và các biến số. HS cũng cần nghiên cứu yêu cầu bài toán để hiểu rõ những mục tiêu phân tích cần đạt được.

Bước 2: Tìm giải pháp

Trong bước này, HS làm việc theo nhóm hoặc cá nhân để thảo luận và đề xuất các phương pháp phân tích dữ liệu. Xác định các đại lượng thống kê cần được xác định như trung bình, trung vị, phương sai hoặc tỉ lệ phần trăm. GV khuyến khích HS đưa ra các dự đoán hoặc giả thuyết dựa trên hiểu biết ban đầu về dữ liệu. Việc này giúp HS rèn luyện khả năng suy luận và phát triển kỹ năng giải quyết vấn đề.

Bước 3: Phân tích dữ liệu

HS sử dụng phần mềm giáo dục thống kê để thực hiện các thao tác phân tích dữ liệu. Các hoạt động bao gồm xử lý dữ liệu, tạo biểu đồ cần thiết (biểu đồ hình quạt tròn, biểu đồ tranh, biểu đồ dạng cột, biểu đồ đoạn thẳng), tính toán các đại lượng thống kê và khám phá các xu hướng hoặc mối quan hệ trong dữ liệu.

Bước 4: Kết luận và vận dụng

HS trình bày kết quả phân tích thông qua báo cáo hoặc thuyết trình, trả lời các câu hỏi đã được đặt ra. HS cần liên hệ các kết quả với thực tiễn, giải thích ý nghĩa và ứng dụng của các kết quả trong ngữ cảnh thực tế. Việc này không chỉ giúp HS tổng hợp và khái quát hóa những gì đã học mà còn phát triển khả năng vận dụng kiến thức thống kê vào giải quyết các vấn đề thực tiễn.

Tổ chức hoạt động khám phá dữ liệu thực tế cho học sinh ở trường Trung học phổ thông với sự hỗ trợ...

Hoạt động khám phá dữ liệu thực tế giúp HS phát triển khả năng suy luận thống kê và tư duy thống kê một cách hiệu quả. Thông qua các nhiệm vụ phân tích dữ liệu, HS nhận thức rõ vai trò của bối cảnh trong việc hiểu và xử lý dữ liệu. Việc sử dụng phần mềm giáo dục thống kê không chỉ hỗ trợ HS xác định các xu hướng và mối quan hệ trong dữ liệu mà còn giúp họ xây dựng lập luận logic dựa trên bằng chứng xác thực. Bên cạnh đó, quá trình trình bày kết quả và liên hệ với thực tiễn tạo điều kiện để HS rèn luyện kỹ năng tổng hợp, diễn giải và đưa ra các quyết định dựa trên cơ sở số liệu một cách khoa học.

2.4. Minh họa hoạt động khám phá bộ dữ liệu về thông tin của học sinh

Hoạt động khám phá dữ liệu sau đây được thiết kế phù hợp với HS từ lớp 10, sau khi HS đã học xong chủ đề các số đặc trưng của mẫu số liệu không ghép nhóm.

**** Giai đoạn 1: Chuẩn bị***

(CB-1) GV cho HS làm quen với phần mềm CODAP theo những gợi ý ở mục 2.3. (CB-2) Dữ liệu minh họa được khai thác từ Kaggle [15], bao gồm thông tin thu thập thông qua một cuộc khảo sát đối với HS trung học tại Bồ Đào Nha. Dữ liệu được thu thập với mục đích dự đoán kết quả học tập của HS dựa trên các yếu tố liên quan, chẳng hạn như thông tin cá nhân, đặc điểm gia đình và môi trường học tập. Bộ dữ liệu không chỉ cung cấp nguồn thông tin phong phú cho việc phân tích mà còn minh họa rõ ràng cách các công cụ thống kê và phần mềm hỗ trợ có thể được áp dụng để giải quyết các vấn đề thực tiễn trong giáo dục. Bộ dữ liệu đã được lược bỏ một số trường thông tin để phù hợp với đối tượng và mục đích học tập. Các thông tin chính gồm có: Giới tính; Thời gian di chuyển đến trường; Thời gian tự học trong tuần; Có hay không sử dụng mạng Internet tại gia đình; Có đang trong một mối quan hệ tình cảm lãng mạn hay không; Tình trạng sức khỏe hiện tại; Số ngày đã nghỉ học trong năm học; Điểm thi Toán (thang điểm 20).

Từ bộ dữ liệu trên, các câu hỏi được thiết kế nhằm hướng dẫn HS khám phá hành vi của nhóm HS tham gia khảo sát, đồng thời tìm hiểu các yếu tố ảnh hưởng đến kết quả điểm thi môn Toán. Những câu hỏi này không chỉ giúp HS phân tích dữ liệu một cách có hệ thống mà còn khuyến khích suy luận thống kê thông qua việc kết nối dữ liệu với bối cảnh thực tế. Các câu hỏi cụ thể được trình bày chi tiết trong Bảng 2.

**** Giai đoạn 2: Tổ chức hoạt động khám phá dữ liệu***

Bước 1: Tìm hiểu vấn đề

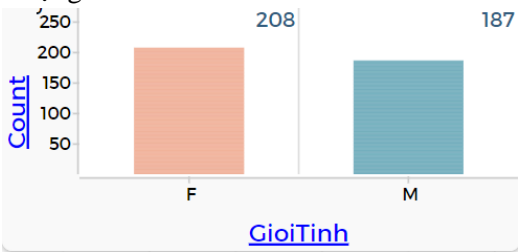
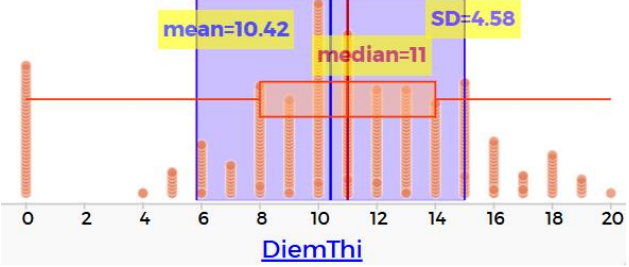
HS tiến hành tìm hiểu bối cảnh liên quan đến bộ dữ liệu, bao gồm đối tượng nghiên cứu và mục đích phân tích, nhằm khám phá các yếu tố tác động đến kết quả học tập và hành vi của HS. Trong quá trình này, HS được làm quen với bộ dữ liệu thông qua việc xem xét các biến số, phân tích loại dữ liệu tương ứng (chẳng hạn như dữ liệu định tính, định lượng, rời rạc hoặc liên tục). Đồng thời, HS cần nghiên cứu kỹ các yêu cầu được đưa ra trong đề bài để hiểu rõ mục tiêu và định hướng của nhiệm vụ phân tích.

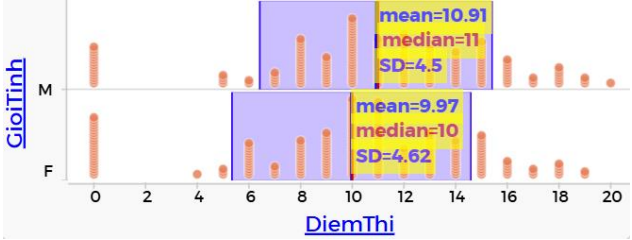
Bước 2: Tìm giải pháp

HS tham gia thảo luận để xác định các đại lượng thống kê cần tính toán nhằm đưa ra câu trả lời cho các câu hỏi. GV khuyến khích HS đưa ra dự đoán trước khi tiến hành phân tích và giải thích các mối quan hệ tiềm năng trong dữ liệu. Chẳng hạn, HS có thể suy luận về mối liên hệ giữa việc sử dụng Internet tại nhà và số ngày nghỉ học hoặc đánh giá tác động của thời gian di chuyển đến trường đối với điểm thi môn Toán. Việc này giúp HS phát triển khả năng lập luận dựa trên dữ liệu và kiểm chứng các giả thuyết đặt ra.

Bước 3: Phân tích dữ liệu

Bảng 2. Câu hỏi và dự kiến kết quả phân tích dữ liệu

Câu hỏi	Dự kiến kết quả phân tích
1. Số lượng HS đã trả lời khảo sát?	Quan sát số dòng của bảng tính, có 395 HS.
2. Tỷ lệ HS nam, nữ tham gia khảo sát là bao nhiêu?	<p>Sử dụng công cụ đồ thị (Graph) trên CODAP để hiển thị số lượng đối tượng theo biến <i>GioiTinh</i>:</p>  <p>187 HS nam, tỉ lệ 47,3% 208 HS nữ, tỉ lệ 52,7%</p>
3. Xác định các số đo xu hướng trung tâm và độ phân tán của mẫu số liệu điểm thi. Những đại lượng này cho biết điều gì về điểm thi của các HS?	 <p>Lựa chọn hiển thị các đại lượng cần xác định trên công cụ đồ thị (Graph).</p> <p>Trung bình $\bar{x} = 10,42$, trung vị $M_e = 11$</p> <p>Tứ phân vị: $Q_1 = 8, Q_2 = 11, Q_3 = 14$</p> <p>Độ lệch chuẩn $SD = 4,58$, Khoảng tứ phân vị $\Delta Q = 6$</p> <p>Khoảng biến thiên là 20.</p> <p>Điểm trung bình (10,42) thấp hơn điểm trung vị (11), điều này cho thấy phân bố điểm bị lệch trái. Nghĩa là có một số lượng đáng kể HS đạt điểm thấp, kéo điểm trung bình xuống. Điều này được củng cố thêm bởi số lượng HS đạt 0 điểm (38 HS), chiếm gần 10% tổng số HS.</p> <p>Độ lệch chuẩn 4,58 khá lớn, cho thấy sự phân tán mạnh của điểm số. Khoảng tứ phân vị 6 điểm cũng chứng minh sự khác biệt đáng kể giữa các nhóm HS.</p>
4. Xác định các giá trị bất thường của điểm thi (nếu có)?	<p>Chọn chức năng hiển thị giá trị bất thường trên CODAP (Show Outliers), phần mềm cho thấy không có giá trị bất thường.</p> <p>HS giải thích dựa vào kiến thức được học, tất cả điểm thi đều thuộc đoạn $[Q_1 - 1,5\Delta Q; Q_3 + 1,5\Delta Q]$.</p>

<p>5. Có sự khác biệt về điểm thi giữa HS nam và HS nữ hay không?</p>	 <p>Phân tích điểm thi theo giới tính, thu được kết quả:</p> <p>HS nam: $\bar{x} = 10,91$, $M_e = 11$, $SD = 4,5$</p> <p>HS nữ: $\bar{x} = 9,97$, $M_e = 10$, $SD = 4,62$</p> <p>Điểm trung bình và trung vị của HS nam cao hơn cho thấy nam sinh có xu hướng đạt điểm cao hơn nữ sinh trong kì thi này. Tuy nhiên, sự khác biệt không quá lớn.</p>
<p>6. So sánh điểm thi giữa các nhóm HS phân chia theo: Thời gian tự học; Thời gian di chuyển đến trường; Tình trạng sức khỏe; Việc sử dụng Internet.</p>	<p>Thực hiện phân tích trên phần mềm cho thấy:</p> <ul style="list-style-type: none"> + Điểm thi có xu hướng tăng lên khi thời gian tự học tăng. + Điểm thi có xu hướng giảm khi thời gian di chuyển tăng. + Khi chia tình trạng sức khỏe thành 5 mức giảm dần từ rất tốt đến rất tệ, nhóm HS với sức khỏe rất tệ có điểm thi cao nhất, sau đó đến nhóm rất tốt. Không có sự khác biệt rõ rệt giữa 3 nhóm còn lại. + Nhóm sử dụng Internet có điểm thi cao hơn nhóm không sử dụng.

Bước 4: Kết luận và vận dụng

- + HS trình bày kết quả phân tích và trả lời câu hỏi. So sánh với dự đoán ban đầu, giải thích.
- + GV đặt câu hỏi để HS rút ra kết luận, ý nghĩa từ kết quả tìm được. Giải thích nguyên nhân có sự khác biệt điểm thi giữa các nhóm. Ví dụ: Vì sao điểm thi có xu hướng giảm khi thời gian di chuyển đến trường tăng; Lí do nào có thể dẫn đến việc nhóm sức khỏe rất tệ có điểm thi cao hơn hẳn các nhóm còn lại?
- + Khuyến khích HS đặt thêm các câu hỏi khác có thể khám phá từ bộ dữ liệu. Ví dụ: Số ngày nghỉ học ảnh hưởng đến điểm thi như thế nào? So sánh số ngày nghỉ học của HS khi phân chia theo các tiêu chí khác nhau như tình trạng sức khỏe, thời gian di chuyển.
- + Từ việc khám phá bộ dữ liệu trên, khuyến khích HS liên hệ với bản thân, đưa ra kế hoạch để đạt được kết quả học tập cao hơn.

3. Kết luận

Việc tổ chức hoạt động khám phá dữ liệu thực tế trong dạy học thống kê tại trường Trung học phổ thông không chỉ giúp học sinh hiểu sâu hơn các khái niệm thống kê mà còn phát triển khả năng suy luận và vận dụng kiến thức vào thực tiễn. Bài báo đã giới thiệu một số phần mềm giáo dục thống kê, trong đó CODAP được sử dụng như một công cụ hỗ trợ hiệu quả đồng thời đề xuất quy trình tổ chức hoạt động khám phá dữ liệu thực tế gồm hai giai đoạn là giai đoạn Chuẩn bị và giai đoạn Tổ chức hoạt động khám phá dữ liệu. Bài báo cũng đưa ra các ví dụ minh họa cụ thể về hoạt động khám phá bộ dữ liệu về thông tin của học sinh nhằm cung cấp những gợi ý thiết thực cho GV trong việc đổi mới phương pháp dạy học môn Toán.

Công nghệ thông tin đã và đang trở thành một phần không thể thiếu trong giáo dục nói chung và dạy học thống kê nói riêng. Các công cụ như CODAP không chỉ giúp HS tiếp cận với dữ liệu thực tế mà còn thúc đẩy tư duy thống kê, giúp HS hình thành kỹ năng giải quyết vấn đề dựa trên dữ liệu. Tuy nhiên, việc ứng dụng CNTT trong giảng dạy cần được thực hiện một cách có chiến lược và cân nhắc để đảm bảo hiệu quả và phù hợp với điều kiện thực tiễn. Cần đặc biệt chú ý rằng mục đích của việc sử dụng công nghệ trong dạy học thống kê không phải là đào tạo HS trở thành người sử dụng phần mềm thống kê, mà là phát triển khả năng suy luận thống kê và tư duy thống kê. GV cần thiết kế các hoạt động sao cho phần mềm đóng vai trò hỗ trợ, trong khi HS tập trung vào việc hiểu, phân tích và đưa ra kết luận dựa trên dữ liệu thực tế.

TÀI LIỆU THAM KHẢO

- [1] Cobb GW & Moore DS, (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly*, 104(9), 801-823.
- [2] T Cường, TN Thắng, J Phouthavong, (2022). Xây dựng kịch bản dạy học với học liệu số phần hình học giải tích theo chương trình môn Toán lớp 10 của nước Cộng hòa Dân chủ Nhân dân Lào. *HNUE Journal of Science*, 67(5), 179-192.
- [3] Bộ Giáo dục & Đào tạo, (2018). *Chương trình giáo dục phổ thông, Chương trình tổng thể* (Ban hành kèm theo Thông tư số 32/2018/TT-BGDĐT ngày 26 tháng 12 năm 2018 của Bộ trưởng Bộ Giáo dục và Đào tạo).
- [4] Ben-Zvi D, (2004). Reasoning about data analysis, in *The challenge of developing statistical literacy, reasoning and thinking*. Springer, p. 121-145.
- [5] Biehler R, (2013). Technological for enhancing statistical reasoning at the school level, in *Third International Handbook of Mathematics Education*, p. 643-689.
- [6] Pratt D, Davies N & Connor D, (2011). The role of technology in teaching and learning statistics, in *Teaching statistics in school mathematics-challenges for teaching and teacher education: A Joint ICMI/IASE Study: The 18th ICMI Study*. Springer, 97-107.
- [7] Wild CJ & Pfannkuch M, (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67(3), 223-248.
- [8] Cobb GW, (2007). One possible frame for thinking about experiential learning. *International Statistical Review*, 75(3), 336-347.
- [9] del Mas RC, (2004). A comparison of mathematical and statistical reasoning, in *The challenge of developing statistical literacy, reasoning and thinking*. Springer, p. 79-95.
- [10] Garfield J & Ben-Zvi D, (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International statistical review*, 75(3), 372-396.
- [11] Gould R, (2010). Statistics and the Modern Student. *International Statistical Review*, 78(2), 297-315.
- [12] Moore DS, (1997). New pedagogy and new content: The case of statistics. *International statistical review*, 65(2), 123-137.
- [13] Biehler R, (1997). Software for learning and for doing statistics. *International Statistical Review*, 65(2), 167-189.
- [14] Pratt D, Davies N & Connor D, (2011). The role of technology in teaching and learning statistics, in *Teaching statistics in school mathematics-challenges for teaching and teacher education*. Springer, p. 97-107.
- [15] <https://www.kaggle.com/datasets/uciml/student-alcohol-consumption>.