

Bài báo nghiên cứu

**DẠY HỌC THỐNG KÊ: NGHIÊN CỨU SO SÁNH GIỮA
CHƯƠNG TRÌNH GIÁO DỤC PHỔ THÔNG
VÀ CHƯƠNG TRÌNH ĐÀO TẠO SINH VIÊN SƯ PHẠM***Nguyễn Thị Nga*, Trần Nguyễn Khánh**Trường Đại học Sư phạm Thành phố Hồ Chí Minh, Việt Nam***Tác giả liên hệ: Nguyễn Thị Nga – Email: ngant@hcmue.edu.vn**Ngày nhận bài: 28-4-2020; ngày nhận bài sửa: 07-5-2020; ngày duyệt đăng: 24-8-2020***TÓM TẮT**

Trong Chương trình (CT) giáo dục phổ thông môn Toán 2018, nội dung thống kê có nhiều sự thay đổi và được giảng dạy xuyên suốt từ lớp 2 đến lớp 12, thay vì chỉ được giảng dạy ở lớp 7 và lớp 10 như CT năm 2006. Bài báo trình bày những kết quả nghiên cứu nội dung thống kê ở CT 2018, so sánh đối chiếu với CT 2006 và chương trình đào tạo sinh viên (CTĐT SV) của Khoa Toán – Tin học, Trường Đại học Sư phạm Thành phố Hồ Chí Minh. Kết quả cho thấy một số nội dung có đề cập trong CT 2018 nhưng lại chưa xuất hiện trong CTĐT SV năm 2016. Từ đó, chúng tôi xây dựng một bài khảo sát trên đối tượng SV năm 3 và năm 4 Khoa Toán – Tin học để tìm hiểu quan niệm của họ về những nội dung thống kê mới được đề cập trong chương trình môn Toán 2018 và những khó khăn mà họ gặp phải. Sau cùng, chúng tôi dựa trên những khó khăn mà SV gặp phải, đề xuất một số giải pháp khắc phục.

Từ khóa: dạy học thống kê; Chương trình môn Toán 2018; Thống kê mô tả

1. Giới thiệu

Thống kê là một môn khoa học có tính thực tiễn lớn, có ứng dụng trong nhiều ngành nghề, lĩnh vực trong cuộc sống. Đặc biệt, nó có ý nghĩa quan trọng đối với các ngành kinh tế. Nó được sử dụng để hiểu hệ thống đo lường biến động, kiểm soát quá trình, cho dữ liệu tóm tắt, và có cơ sở đưa ra quyết định dựa trên dữ liệu.

Ngày 26 tháng 12 năm 2018, Bộ Giáo dục và Đào tạo đã ban hành CT giáo dục phổ thông tổng thể kèm với khung CT tổng thể và khung CT các môn học. CT môn Toán 2018 đã chỉ rõ vai trò của mạch kiến thức Thống kê và Xác suất như sau:

Thống kê và Xác suất là một thành phần bắt buộc của giáo dục toán học trong nhà trường, góp phần tăng cường tính ứng dụng và giá trị thiết thực của giáo dục toán học. Thống kê và Xác suất cung cấp cho HS công cụ lý thuyết để phân tích các dữ liệu thông tin thể hiện dưới nhiều hình thức khác nhau, hiểu được bản chất xác suất, hình thành sự hiểu biết về vai trò của thống kê như là một nguồn thông tin quan trọng về mặt xã hội. Từ đó, nâng cao sự hiểu biết và phương pháp nghiên cứu các hiện tượng xã hội của thế giới hiện đại cho học sinh. (Ministry of Education and Training, 2018)

Cite this article as: Nguyen Thi Nga, & Tran Nguyen Khanh (2020). Statistics teaching: A comparison study between high school programme and teacher training program. *Ho Chi Minh City University of Education Journal of Science*, 17(8), 1398-1409.

Vậy liên quan đến nội dung Thống kê, CT môn Toán 2018 có những điểm gì mới so với CT môn Toán 2006? Nội dung thống kê được giảng dạy ở Khoa Toán – Tin học, Trường Đại học Sư phạm Thành phố Hồ Chí Minh có đáp ứng được sự thay đổi này hay không? Sinh viên đã có những hiểu biết gì về những nội dung thống kê mới được đề cập đến trong CT môn Toán 2018?

2. So sánh nội dung thống kê trong chương trình môn Toán 2018, chương trình môn Toán 2006 và chương trình đào tạo sinh viên

Trong chương trình môn Toán 2006 (chương trình hiện hành), nội dung Thống kê được giảng dạy ở lớp 7 và lớp 10. Vì vậy, chúng tôi sẽ tiến hành nghiên cứu CT 2006 thông qua sách giáo khoa (SGK) Toán 7, Đại số 10 và Đại số 10 nâng cao để so sánh với CT giáo dục phổ thông môn Toán 2018. Ngoài ra, chúng tôi cũng phân tích so sánh với nội dung Thống kê trong giáo trình “Xác suất thống kê và quá trình ngẫu nhiên” của tác giả Nguyễn Chí Long (2008), Nhà xuất bản Đại học Quốc gia Thành phố Hồ Chí Minh. Đây là giáo trình được sử dụng trong học phần Xác suất thống kê 1 ở Khoa Toán – Tin học Trường Đại học Sư phạm Thành phố Hồ Chí Minh.

Từ các kết quả phân tích, chúng tôi so sánh các CT dựa trên hai nội dung chính: Thu thập và tổ chức dữ liệu; Phân tích và xử lý số liệu.

2.1. Thu thập và tổ chức dữ liệu

Bảng 1. So sánh nội dung thu thập số liệu giữa CT 2018, CT 2006 và CTĐTSV

	CT 2018	CT 2006	CTĐTSV
Thu thập số liệu	Chú trọng vào việc thu thập số liệu thông qua 2 yêu cầu cần đạt chính: - Thực hiện và lí giải được việc thu thập, phân loại dữ liệu theo các tiêu chí cho trước từ nhiều nguồn gốc khác nhau: văn bản; bảng kiểu, kiến thức trong các lĩnh vực giáo dục khác; phỏng vấn, truyền thông, Internet; thực tiễn - Nhận biết, chứng tỏ được tính hợp lí của dữ liệu theo các tiêu chí toán học đơn giản	Không có hoạt động nào để giúp học sinh phát triển, rèn luyện khả năng thu thập số liệu. Chú trọng vào việc tổ chức dữ liệu hơn là thu thập số liệu	Chỉ đề cập đến các phương pháp chọn mẫu trong thống kê

Bình luận: CT 2018 đặc biệt chú trọng vào tính thực tiễn. Việc thu thập dữ liệu là một phần quan trọng, do đó CT 2018 chú trọng vào ý nghĩa của việc thu thập dữ liệu. Nội dung thu thập dữ liệu không được CT 2006 quan tâm đến. Bên cạnh đó CTĐTSV không đề cập thu thập số liệu mà chỉ đề cập các phương pháp chọn mẫu trong thống kê dưới góc độ lí thuyết.

Bảng 2. So sánh nội dung đọc và mô tả số liệu giữa CT 2018, CT 2006 và CTĐTSV

	CT 2018	CT 2006	CTĐTSV
Đọc và mô tả số liệu	Ngoài những nội dung như CT 2006, CT 2018 còn đề cập biểu đồ đoạn thẳng. Ngoài việc chú trọng thu gọn dữ liệu và vẽ biểu đồ, Chương trình 2018 còn chú trọng đến việc: - Lựa chọn dạng biểu diễn dữ liệu sao cho phù hợp - Phát hiện vấn đề hoặc quy luật đơn giản dựa trên việc quan sát bảng số liệu, hay biểu đồ - Nhận biết các số liệu không chính xác	Đề cập hai vấn đề: - Kỹ thuật lập bảng tần số, tần suất; bảng tần số, tần suất ghép lớp - Kỹ thuật vẽ các biểu đồ: Hình cột, hình quạt, Đường gấp khúc tần số, tần suất	Đề cập: - Bảng phân phối tần số, tần suất. - Đường gấp khúc phân phối và Biểu đồ tổ chức

Bình luận: Trong CT 2018, có nhiều yêu cầu cần đạt gắn với thực tiễn. Cụ thể, CT nhấn mạnh ý nghĩa của việc thu thập và tổ chức dữ liệu trong thực tiễn, gắn kết với các môn học khác. Trong khi đó CT 2006 chỉ đơn thuần đưa ra các nhiệm vụ nhằm hướng đến kỹ thuật tính toán, vẽ biểu đồ với yêu cầu cho trước. Hay CTĐTSSV chỉ đề cập đến bảng phân phối thực nghiệm, đường gấp khúc phân phối và biểu đồ tổ chức. Tuy nhiên về ý nghĩa của việc biểu diễn phân bố dữ liệu hay so sánh hai mẫu số liệu thì không được đề cập đến.

2.2. Phân tích và xử lý số liệu

Bảng 3. So sánh nội dung các tham số định tâm giữa CT 2018, CT 2006 và CTĐTSSV

	CT 2018	CT 2006	CTĐTSSV
Các tham số định tâm	<ul style="list-style-type: none"> - Xuất hiện tham số mới: Tứ phân vị - Ngoài việc chú trọng vào kỹ thuật tính các giá trị, CT còn tập trung vào việc giải quyết vấn đề thực tiễn thông qua ý nghĩa của từng tham số thống kê 	<ul style="list-style-type: none"> - Kiểu nhiệm vụ tính số trung bình cộng chiếm ưu thế hơn hẳn các kiểu nhiệm vụ khác. Không đưa ra nhiệm vụ cho HS chọn số trung bình, trung vị hay một làm đại diện cho mẫu số liệu mà hầu như chỉ yêu cầu HS tính toán các giá trị - Không đặt HS vào tình huống thực tiễn cần vận dụng ý nghĩa của các tham số định tâm 	<ul style="list-style-type: none"> - Số trung vị, mốt không được nhắc đến trong chương trình - Đưa ra các công thức tính toán trong nhiều trường hợp - Không đưa ra cách tính, công thức trong trường hợp các dữ liệu ghép lớp - Ý nghĩa của từng công thức, từng khái niệm không được nhắc đến

Bình luận: CT 2018 chú trọng ý nghĩa của các tham số định tâm và vận dụng chúng để đưa ra những kết luận có giá trị trong tình huống thực tiễn. CT 2006 hay CTĐTSSV chỉ đề cập những công thức tính toán mà không có tình huống khai thác đến ý nghĩa của các tham số. Ngoài ra, nội dung các giá trị tứ phân vị không có mặt trong cả CT 2006 và CTĐTSSV.

Bảng 4. So sánh nội dung các tham số đo độ phân tán giữa CT 2018, CT 2006 và CTĐTSSV

	CT 2018	CT 2006	CTĐTSSV
Các tham số đo độ phân tán	<ul style="list-style-type: none"> - Ngoài phương sai và độ lệch chuẩn, CT còn đề cập đến hai khái niệm mới là biên độ (khoảng biến thiên) và khoảng tứ phân vị - Chú trọng đến các tham số đo độ phân tán đối với mẫu số liệu ghép lớp - Chú ý đến ý nghĩa của từng tham số và việc đưa ra nhận xét, bình luận dựa vào ý nghĩa của từng tham số trong từng trường hợp 	<ul style="list-style-type: none"> Chỉ đề cập hai giá trị phương sai và độ lệch chuẩn. Kỹ thuật tính toán và ý nghĩa của chúng được đề cập đầy đủ. Tuy nhiên, CT chỉ chú trọng đến kỹ năng tính toán, việc khai thác các kết quả từ việc tính toán ấy thì không được chú trọng 	<ul style="list-style-type: none"> - Đưa ra các công thức tính toán trong nhiều trường hợp, tuy nhiên về mặt ý nghĩa của từng công thức, từng khái niệm thì không được nhắc đến - Không đưa ra cách tính, công thức trong trường hợp các dữ liệu ghép lớp - Không đưa vào tình huống thực tiễn cần vận dụng ý nghĩa của các tham số

Bình luận: Trong CT 2018, yêu cầu cần đạt là HS có thể đưa ra những đánh giá có cơ sở và hợp lý khi giải quyết một vấn đề thực tiễn. Ngoài ra khoảng tứ phân vị có thể giúp HS nhận ra được những giá trị ngoại lai mà bằng biểu đồ hay số liệu ta rất khó phân biệt, từ đó có thể đưa ra quy luật cho dữ liệu một cách chính xác nhất.

Kết luận: Từ những phân tích trên cho thấy nội dung thống kê ở CT môn Toán 2018 có nhiều sự thay đổi khác biệt. Ngoài việc nắm được các định nghĩa, kĩ thuật tính toán, HS cần phải giải quyết được các vấn đề dựa trên khai thác ý nghĩa của từng tham số thống kê. Tuy nhiên, CTĐTSTV ở học phần Xác suất thống kê 1 không cung cấp ý nghĩa của từng tham số thống kê. Ngoài ra, CT môn Toán 2006 cũng hạn chế ở việc khai thác ý nghĩa của các tham số. Như vậy, với nhiệm vụ dạy học ý nghĩa của các tham số thống kê trong CT 2018, SV Khoa Toán – Tin học sẽ phải gặp phải những khó khăn gì? Mối quan hệ cá nhân của sinh viên đối với những nội dung mới trong CT 2018 như thế nào?

3. Nghiên cứu thực nghiệm

Để trả lời được các câu hỏi trên, chúng tôi đã thiết kế bốn bài toán khai thác đến những khác biệt của chương trình môn Toán 2018 và tổ chức điều tra trên 80 SV năm 3 và năm 4 của Khoa Toán – Tin học, Trường Đại học Sư phạm Thành Phố Hồ Chí Minh.

3.1. Giới thiệu bài toán

Bài toán 1. Để có thể đưa ra câu trả lời cho câu hỏi “Ở Thành phố Hồ Chí Minh, liệu rằng con cái thường cao hơn bố mẹ hay không?” người ta thực hiện các cuộc khảo sát trên các gia đình. Trong đó, có những nhóm thực hiện khảo sát như sau:

Nhóm 1:

- Thời gian khảo sát: 20/12/2019
- Đối tượng khảo sát: 1000 hộ gia đình có con trên 18 tuổi trên các địa bàn quận trung tâm Thành phố Hồ Chí Minh: Quận 1, Quận 3, Quận 5, Quận 10, Quận Bình Thạnh, Quận Phú Nhuận.
- Thông tin khảo sát: giới tính, chiều cao.

Nhóm 2:

- Thời gian khảo sát: 20/12/2019
- Đối tượng khảo sát: 1500 hộ gia đình trên địa bàn các quận trung tâm Thành phố Hồ Chí Minh: Quận 1, Quận 3, Quận 5, Quận 10, Quận Bình Thạnh, Quận Phú Nhuận.
- Thông tin khảo sát: Giới tính, chiều cao, cân nặng.

Nhóm 3:

- Thời gian khảo sát: 20/12/2019
- Đối tượng khảo sát: 1000 hộ gia đình có con từ 18 tuổi trở lên trên tất cả các quận, huyện trên địa bàn TP. Hồ Chí Minh.
- Thông tin khảo sát: giới tính, chiều cao.

Nhóm 4:

- Thời gian khảo sát: 05/01/2020.
- Đối tượng khảo sát: 1500 hộ gia đình trên tất cả các quận, huyện trên địa bàn Thành phố Hồ Chí Minh.
- Thông tin khảo sát: giới tính, chiều cao, cân nặng

Theo bạn, kết quả thống kê của nhóm nào sẽ đưa ra câu trả lời chính xác nhất? Hãy giải thích tại sao?

Câu trả lời được mong đợi:

Trả lời 1 (TL1): Lựa chọn dựa trên cỡ mẫu, đảm bảo tính đại diện và thông tin cần nghiên cứu.

Nhóm 3 đảm bảo đầy đủ các yếu tố để phục vụ cho nghiên cứu: Cỡ mẫu lớn, mẫu phải có đầy đủ thành phần tạo nên tổng thể, thông tin điều tra phải phù hợp với vấn đề cần nghiên cứu (So sánh chiều

cao của bố mẹ với con cái, thông tin ở đây là độ tuổi con phải từ 18 tuổi trở lên mới có cơ sở để so sánh bởi vì ở độ tuổi này về cơ bản đã phát triển hoàn thiện về chiều cao).

Để giải quyết Bài toán 1 có thể còn có các trả lời sau:

Trả lời 2 (TL2): Lựa chọn dựa trên “sự nổi trội”

- SV có thể quan tâm đến sự phát triển và nổi trội của các vùng. SV chỉ rút gọn điều tra trên các địa phương phát triển nổi trội và cho rằng có thể đại diện được cho tổng thể. Kết hợp với cỡ mẫu lớn và tính nổi trội thì SV có thể chọn phương án của nhóm 2.

Trả lời 3 (TL3): So sánh cỡ mẫu

SV sẽ chọn nhóm có cỡ mẫu cao nhất trong số 4 nhóm – nhóm 4. Ở nhóm 4 đã đưa ra tiêu chí chọn mẫu khá hợp lí: cỡ mẫu lớn, điều tra đảm bảo tính đại diện cho tổng thể. Tuy nhiên ở nhóm 4, không hề đề cập độ tuổi của con cái mặc dù độ tuổi của con cái là một thông tin rất quan trọng trong cuộc điều tra này.

Bài toán 2. (tham khảo Nguyễn Ngọc Đan (2017)) Tại phòng khám đa khoa X, người ta ghi lại số ngày khỏi bệnh B sau khi dùng thuốc theo đơn bác sĩ của 11 bệnh nhân như sau: {1; 1; 1; 2; 2; 2; 3; 3; 3; 3; 3; 21}

a) Một bệnh nhân đến khám hỏi rằng họ có thể khỏi bệnh sau khoảng bao nhiêu ngày. Nếu là bác sĩ, bạn sẽ trả lời là bao nhiêu ngày? Tại sao bạn chọn con số đó để trả lời?

b) Nếu bạn được hỏi rằng, đa số bệnh nhân khỏi bệnh sau bao nhiêu ngày, bạn sẽ trả lời thế nào? Vì sao?

c) Giả sử 1 ngày điều trị, bảo hiểm y tế phải chi trả 100 nghìn đồng. Nếu bạn là nhân viên công ti BHYT, dựa vào các số liệu trên bạn sẽ dự trù chi phí điều trị cho một bệnh nhân là bao nhiêu?

Câu trả lời được chúng tôi mong đợi:

Trả lời 1 (TL1): Số trung vị

Đây là câu trả lời được mong đợi cho ý a của bài toán. Bởi vì bài toán này bị ảnh hưởng bởi giá trị ngoại lai là 21. Và nếu sử dụng trả lời số trung bình thì kết quả sẽ ra là sau 4 ngày bệnh nhân sẽ khỏi bệnh. Tuy nhiên ta lại thấy rằng, thực tế không có bệnh nhân nào khỏi sau 4 ngày, mà hơn 50% bệnh nhân khỏe trước 2 ngày (ý nghĩa của số trung vị). Do đó câu trả lời 2 ngày là phù hợp.

Trả lời 2 (TL2): Mode

Đây là câu trả lời được mong đợi cho ý b của bài toán. Trả lời này dễ dàng đạt được bởi vì cụm từ “đa số” khiến người đọc nghĩ ngay là cần phải tìm giá trị xuất hiện nhiều nhất hay là đi tìm mode. Theo dự kiến, đa số SV đưa ra câu trả lời cho ý b của bài toán bằng trả lời này.

Trả lời 3 (TL3): Số trung bình

Đây là câu trả lời được mong đợi cho ý c của bài toán bởi vì đặc tính quan trọng của số trung bình: “san bằng” chênh lệch giữa các giá trị. Tuy nhiên có thể trả lời này được nhiều SV sử dụng để giải quyết ý a và c của bài toán.

Ngoài những câu trả lời trên, SV còn có thể trả lời như sau:

Trả lời 4 (TL4): Giá trị nhỏ nhất, giá trị lớn nhất

SV sẽ dùng Giá trị nhỏ nhất hoặc giá trị lớn nhất để trả lời cho các câu hỏi trên.

Bài toán 3. Dưới đây là bảng số liệu thống kê tuổi thọ của 100 bóng đèn của nhà sản xuất A và nhà sản xuất B. Với cùng một mức giá, nếu là bạn thì bạn sẽ lựa chọn bóng đèn của nhà sản xuất nào? Giải thích vì sao?

Bảng 5. Tuổi thọ 100 bóng đèn của nhà sản xuất A (Đơn vị: giờ)

5600	5610	5620	5620	5620	5630	5650	5660	5665	5670
5680	5700	5720	5735	5740	5750	5760	5780	5800	5810
5820	5835	5840	5845	5850	5850	5870	5890	5890	5895
5900	5920	5930	5940	5950	5950	5960	5960	5960	5960
5965	5965	5970	5970	5980	5980	5980	5980	5980	5990
5990	5995	5995	6000	6000	6000	6000	6000	6010	6010
6015	6015	6015	6015	6020	6020	6020	6025	6025	6025
6030	6030	6030	6040	6045	6045	6045	6050	6050	6050
6050	6050	6060	6060	6065	6075	6075	6080	6095	6100
6120	6150	6180	6190	6190	6200	6230	6250	6260	6290

Bảng 6. Tuổi thọ của 100 bóng đèn của nhà sản xuất B (Đơn vị: giờ)

5710	5710	5720	5720	5740	5740	5740	5760	5760	5770
5770	5780	5780	5790	5790	5800	5810	5820	5830	5850
5860	5870	5875	5880	5890	5890	5890	5910	5910	5910
5910	5920	5920	5920	5920	5930	5930	5930	5930	5940
5940	5940	5940	5940	5940	5945	5945	5945	5945	5945
5950	5950	5950	5960	5960	5960	5970	5970	5975	5975
5980	5980	5980	5980	5980	5980	5980	5990	5990	5990
5990	5990	5995	5995	6000	6000	6000	6010	6020	6040
6070	6080	6090	6130	6140	6140	6140	6140	6150	6150
6160	6170	6170	6170	6180	6190	6190	6210	6210	6210

Câu trả lời mong đợi:

Trả lời 1 (TL1): Ghép lớp dữ liệu

SV ghép lớp dữ liệu thành những lớp có biên độ tùy ý (Có thể biên độ là 50). Cụ thể:

Bảng 7. Bảng phân bố tần số ghép lớp tuổi thọ 100 bóng đèn của nhà sản xuất A

Lớp ghép	Giá trị đại diện	Tần số
[5600;5650)	5625	6
[5650;5700)	5675	5
[5700;5750)	5725	4
[5750;5800)	5775	3
[5800;5850)	5825	6
[5850;5900)	5875	6
[5900;5950)	5925	4
[5950;6000)	5975	19
[6000;6050)	6025	24
[6050;6100)	6075	12
[6100;6150)	6125	2
[6150;6200)	6175	4
[6200;6250)	6225	2
[6250;6300)	6275	3

Tính được $\bar{x}_A = 5955, s \approx 158,11$

Bảng 8. Bảng phân bố tần số ghép lớp tuổi thọ 100 bóng đèn của nhà sản xuất B

Lớp ghép	Giá trị đại diện	Tần số
[5600;5650)	5625	0
[5650;5700)	5675	0
[5700;5750)	5725	7
[5750;5800)	5775	8
[5800;5850)	5825	4
[5850;5900)	5875	8
[5900;5950)	5925	23
[5950;6000)	5975	24
[6000;6050)	6025	6
[6050;6100)	6075	3
[6100;6150)	6125	5
[6150;6200)	6175	9
[6200;6250)	6225	3
[6250;6300)	6275	0

Tính được $\bar{x}_B = 5955, s \approx 128,45$

Ta thấy rằng ở hai nhà sản xuất có cùng giá trị trung bình. Tuy nhiên độ lệch chuẩn của các bóng đèn ở nhà sản xuất B lại nhỏ hơn độ lệch chuẩn của các bóng đèn ở nhà sản xuất A. Do đó bóng đèn ở nhà sản xuất B có độ bền phân bố đều hơn ở nhà sản xuất A. Do đó nên mua bóng đèn ở nhà sản xuất B.

Đối với bài toán này, chúng tôi dự kiến còn có những câu trả lời sau:

Trả lời 2 (TL2): Biên độ

Biên độ của nhà sản xuất A lớn hơn biên độ của nhà sản xuất B. Do đó dữ liệu của nhà sản xuất A biến động nhiều hơn nhà sản xuất B. Do đó, chọn mua bóng đèn ở nhà sản xuất B.

Trả lời 3 (TL3): Số trung bình và độ phân tán

SV sẽ lập bảng tần số, tần suất của dữ liệu và tính số trung bình, phương sai/độ lệch chuẩn của dữ liệu. Tuy nhiên, việc làm này rất khó khăn, mất nhiều thời gian vì có quá nhiều giá trị khác nhau.

Trả lời 4 (TL4): Số trung bình

SV chỉ tính số trung bình của các dữ liệu và so sánh với nhau mà không quan tâm đến các tham số đo độ phân tán của dữ liệu.

Bài toán 4. Một vận động viên nữ A tham gia tập luyện chạy cự li 100m. Kết quả sau 20 ngày luyện tập được trình bày theo Bảng 9 dưới đây:

Bảng 9. Thời gian chạy trong 20 ngày của vận động viên A

Ngày	Thời gian chạy của A (giây)
1	10,49
2	10,58
3	10,73
4	10,82
5	10,75
6	10,90
7	10,82
8	10,85
9	10,78

10	10,76
11	10,73
12	10,77
13	10,78
14	10,81
15	10,73
16	11,25
17	10,75
18	10,70
19	10,65
20	20,75

Huấn luyện viên muốn gửi bài báo cáo thành tích bằng việc tính thời gian chạy trung bình của vận động viên A này; tuy nhiên, để bài báo cáo phản ánh tốt nhất khả năng của vận động viên A, huấn luyện viên sẽ bỏ đi những ngày có xuất hiện giá trị đột xuất (giá trị ngoại lai). Theo bạn, huấn luyện viên sẽ bỏ đi giá trị nào trong số 20 giá trị trên? Giải thích vì sao?

Câu trả lời chúng tôi mong đợi:

Trả lời 1 (TL1): Tính được:

$$Q_1 = 10,73, Q_2 = 10,765, Q_3 = 10,82, IQR = 0,09$$

$$Q_1 - 1,5 \times IQR = 10,595, Q_3 + 1,5 \times IQR = 10,955$$

Do đó giá trị ngoại lai (biến đột xuất) là: 10,49 (ngày 1), 10,58 (ngày 2), 11,25 (Ngày 16), 20,75 (Ngày 20). Như vậy huấn luyện viên sẽ bỏ 4 giá trị này để tính thời gian chạy trung bình của vận động viên A.

Ngoài câu trả lời mong đợi này, dự kiến còn có những câu trả lời sau:

Trả lời 2 (TL2): Dự kiến SV sẽ bỏ đi giá trị lớn nhất và giá trị nhỏ nhất trong Bảng 9. Và do đó hai giá trị 10,49 và giá trị 20,75 bị bỏ.

Trả lời 3 (TL3): SV thấy rằng giá trị 20,75 vượt xa các giá trị khác nên bỏ đi duy nhất giá trị này.

Trả lời 4 (TL4): SV bỏ đi hai giá trị là 20,75 và 11,25 vì nhìn vào Bảng 9 ta thấy hầu như thời gian chạy của vận động viên trên đều dưới 11 giây.

3.2. Kết quả thực nghiệm

3.2.1. Bài toán 1

Dưới đây là bảng kết quả cho Bài toán 1 được khảo sát trên 80 SV năm 3 và năm 4 Khoa Toán – Tin học, Trường Đại học Sư phạm Thành phố Hồ Chí Minh:

Bảng 10. Kết quả khảo sát Bài toán 1

TL1	TL2	TL3	Khác	Không TL	Cộng
39	2	33	4	2	80

Từ Bảng 10 và dựa trên câu trả lời của SV, chúng tôi thấy rằng:

- Đa số SV lựa chọn câu trả lời dựa trên cỡ mẫu, tính ngẫu nhiên và tính đại diện của mẫu.
- Việc lựa chọn thông tin điều tra của nhiều SV còn gặp hạn chế. Có thể lí giải ở đây vì SV chưa quan tâm đến thông tin độ tuổi phát triển và đối tượng so sánh. Do đó, những SV này đưa ra câu trả lời chưa được thỏa đáng. Một lí do khác có thể do nhiệm vụ lựa chọn mẫu nào để đại diện tốt là một nhiệm vụ mới lạ đối với SV mà trong CT môn Toán 2006 và CTĐTSPV chưa đề cập.

3.2.2. Bài toán 2

Và dưới đây là bảng thống kê kết quả cho Bài toán 2:

Bảng 11. Kết quả khảo sát Bài toán 2

	TL1	TL2	TL3	TL4	Khác	Không TL
A	8	19	22	15	4	12
B	0	71	4	1	0	4
C	0	21	50	3	0	6

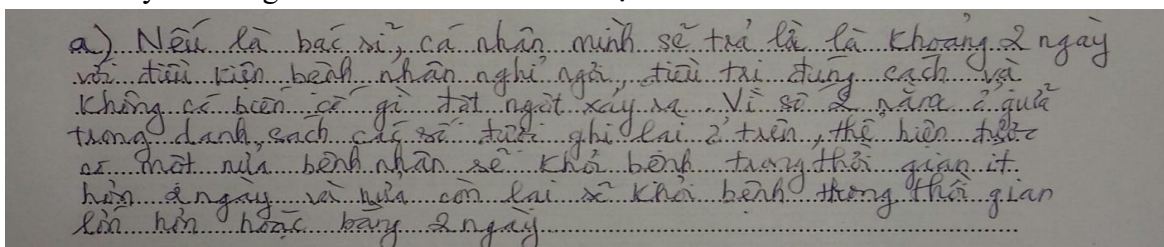
Từ Bảng 11 và phân tích câu trả lời của SV, chúng tôi thu được kết quả sau:

- Đa phần SV không nắm được ý nghĩa của số trung vị là chia đôi mẫu dữ liệu thành hai phần bằng nhau. Nguyên nhân dẫn đến kết quả này là CT môn Toán 2006 chưa chú trọng vào ý nghĩa số trung vị, các bài tập liên quan đến số trung vị còn ít. Các bài tập này chỉ đơn thuần là tính toán số trung vị dựa vào kỹ thuật SGK đã cung cấp, các bài tập khai thác ý nghĩa của số trung vị không xuất hiện. Bên cạnh đó, CTĐTSV không cung cấp đến SV ý nghĩa của từng tham số thống kê, số trung vị và một hoàn toàn vắng bóng trong CTĐTSV.

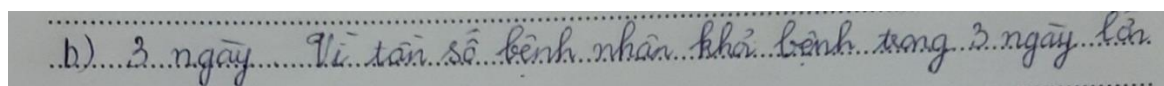
- SV gặp khó khăn khi đứng trước nhiệm vụ lựa chọn tham số thống kê để đại diện tốt cho mẫu. Điều này có thể được lí giải do các kiểu nhiệm vụ khai thác ý nghĩa của từng tham số thống kê thiếu vắng trong CT môn Toán 2006 và CTĐTSV. Các bài tập được nhắc tới đơn thuần chỉ là tính toán trên những công thức và kỹ thuật sẵn có, HS – SV không phải lựa chọn tham số nào đại diện cho mẫu trong từng trường hợp.

- Chỉ có một số ít SV biết được số trung bình chịu ảnh hưởng mạnh mẽ bởi các giá trị ngoại lai (biến đột xuất). Trong những trường hợp mẫu xuất hiện giá trị đột xuất, số trung bình không phải là lựa chọn tốt nhất để đại diện cho mẫu.

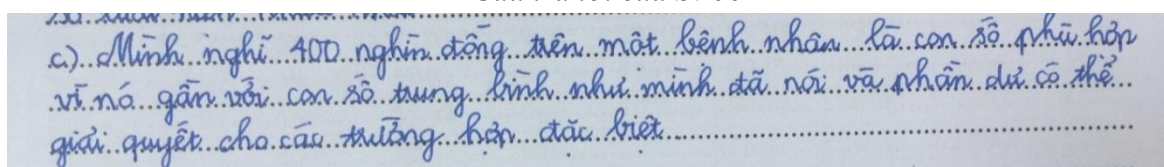
Dưới đây là những câu trả lời chính xác từ một số SV:



Câu trả lời của SV04



Câu trả lời của SV06



Câu trả lời của SV09

3.2.3. Bài toán 3

Đây là bảng thống kê kết quả khảo sát Bài toán 3:

Bảng 12. Kết quả khảo sát Bài toán 3

TL1	TL2	TL3	TL4	Khác	Không TL	Cộng
7	11	0	49	14	0	80

Chỉ có 07/80 SV đưa ra được câu trả lời dựa trên việc ghép lớp và tính toán các tham số thống kê cần thiết.

Từ Bảng 12 và phân tích câu trả lời của SV, chúng tôi đã đạt được những kết quả sau:

- Khi so sánh hai mẫu số liệu, đa số SV chỉ quan tâm đến tham số định tâm (số trung bình) mà không quan tâm đến độ phân tán của dữ liệu. Nguyên nhân chính có thể trong CT môn Toán 2006, số lượng bài tập của kiểu nhiệm vụ nhận xét dựa trên tham số đo độ phân tán còn rất ít. Kiểu nhiệm vụ so sánh hai mẫu số liệu không xuất hiện trong CT. Hơn nữa, ở các bài tập có sự nhận xét về độ phân tán thì SGK luôn hướng dẫn HS tính phương sai, độ lệch chuẩn rồi mới đưa ra nhận xét. Hơn nữa, CTĐTSV không đề cập ý nghĩa của các tham số đo độ phân tán (phương sai, độ lệch chuẩn). Do đó nhiệm vụ so sánh hai mẫu số liệu là một khó khăn đối với SV.

- Đa số SV chưa có kỹ năng xử lý số liệu, cụ thể ở Bài toán 3 là ghép lớp dữ liệu. Đúng trước rất nhiều giá trị khác nhau, SV vẫn tính số trung bình mà không xử lý bảng số liệu, không có ý định ghép lớp để thuận tiện trong việc tính toán các tham số. Nguyên nhân dẫn đến điều này là do đa số các bài tập liên quan đến mẫu số liệu ghép lớp, SGK đều có yêu cầu rõ ràng, và việc lựa chọn độ rộng của lớp ghép cũng đã được SGK đưa ra. Điều này hình thành thói quen cho HS chỉ những bài toán yêu cầu ghép lớp thì HS mới thực hiện. Do đó, ở bài toán này, hầu hết SV vẫn tính toán các giá trị mà không thực hiện ghép lớp mặc dù việc tính toán này là hết sức khó khăn.

- Chỉ có một bộ phận nhỏ SV (8,75%) có kỹ năng xử lý số liệu, đưa ra được kết luận dựa trên số trung bình và tham số đo độ phân tán (cụ thể là độ lệch chuẩn). Điều này chứng tỏ những SV trên nắm được ý nghĩa của việc ghép lớp dữ liệu, ý nghĩa của tham số đo độ phân tán trong việc đưa ra nhận xét dựa trên chúng.

3.2.4. Bài toán 4

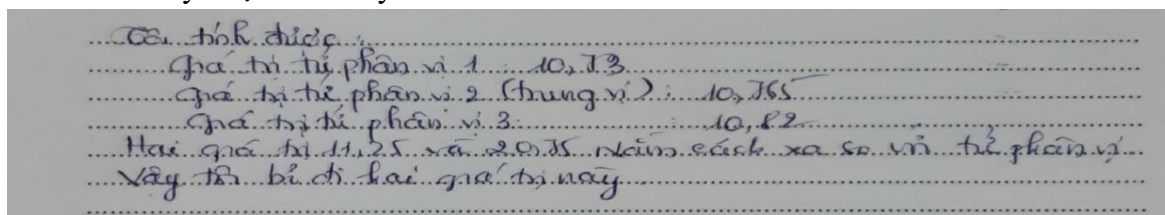
Đây là bảng thống kê kết quả khảo sát của Bài toán 4:

Bảng 13. Kết quả khảo sát Bài toán 4

TL1	TL2	TL3	Không TL	Khác	Cộng
1	41	20	11	7	80

Từ Bảng 13 và câu trả lời của SV, chúng tôi nhận thấy:

- Có duy nhất 1 SV sử dụng Tứ phân vị để trả lời cho câu hỏi trên. Tuy nhiên, câu trả lời của SV này được trình bày như sau:



Câu trả lời của SV01

Mặc dù SV này đã dùng tứ phân vị nhưng vẫn không biết cách tìm ra giá trị ngoại lai thực sự thông qua tứ phân vị.

- Việc ứng dụng của tứ phân vị trong việc tìm ra giá trị ngoại lai là một nhiệm vụ khá xa lạ đối với SV.

- Về cơ bản, SV có thể nhận biết được một số giá trị ngoại lai thông qua dữ liệu thống kê. Tuy nhiên để tìm ra tất cả các giá trị ngoại lai mà không thông qua tứ phân vị thì khó có thể thực hiện được.

Lí do SV gặp phải những khó khăn trên bởi vì tứ phân vị là một khái niệm hoàn toàn mới đối với SV bởi vì trong CT môn Toán 2006 và CTĐTTSV không được nhắc đến.

4. Kết luận

Trong nghiên cứu này, từ việc phân tích CT Toán 2006 (cụ thể qua phân tích SGK Toán lớp 7 và lớp 10 hiện hành) cho thấy nội dung thống kê của CT môn Toán 2006 chưa khai thác ý nghĩa của từng tham số thống kê, việc thu thập số liệu thống kê không được chú trọng. Bên cạnh đó, nội dung thống kê mô tả trong CTĐTTSV xuất hiện rất mờ nhạt. Nội dung thống kê trong CT môn Toán 2018 có nhiều sự thay đổi: chú trọng đến ý nghĩa của việc thu thập và tổ chức dữ liệu; tập trung khai thác ý nghĩa của từng tham số thống kê trong các tình huống thực tiễn; một số tham số, khái niệm mới xuất hiện như biên độ, tứ phân vị. Do đó, với những nội dung thống kê mới xuất hiện ở CT môn Toán 2018, nhiều SV không giải quyết được các bài toán thực tiễn, khai thác ý nghĩa của từng tham số thống kê, vì vậy họ sẽ gặp khó khăn khi giảng dạy. Để khắc phục những khó khăn trên, chúng tôi thiết nghĩ CTĐTTSV nên được bổ sung nội dung cung cấp kiến thức về thống kê mô tả, ý nghĩa của từng tham số thống kê. Đặc biệt, ở nội dung này nên rèn luyện và phát triển năng lực mô hình hóa toán học, cung cấp cho SV một số vấn đề thực tiễn liên quan đến thống kê mô tả. Trong các học phần Lí luận và Phương pháp dạy học môn Toán nên có nhiều các hoạt động trao đổi, thảo luận về những điểm mới trong CT môn Toán 2018, đặc biệt là ở nội dung thống kê. Ngoài ra, nên tăng cường các bài tập thiết kế những tình huống giảng dạy khai thác ý nghĩa của từng tham số thống kê, ý nghĩa của việc thu thập và tổ chức dữ liệu... để sinh viên tập giảng.

❖ **Tuyên bố về quyền lợi:** Các tác giả xác nhận hoàn toàn không có xung đột về quyền lợi.

TÀI LIỆU THAM KHẢO

- Doan, Q. (Chief Editor), Nguyen, H. D. (Editor), Nguyen, X. L., Dang, H. T., & Tran V. L. (2013). *Dại số 10 nâng cao [Advanced Algebra 10]*. Vietnam Education Publishing House.
- Ministry of Education and Training. (December 26th, 2018). *Mathematics Education curriculum*. Retrieved from <https://data.moet.gov.vn/index.php/s/m6ztfi7sUIIGQdY#pdfviewer>
- Nguyen, C. L. (2008). *Xác suất thống kê và Qua trình ngẫu nhiên [Probability Statistics and random processes]*. Vietnam National University, Ho Chi Minh City Publishing House.
- Nguyen, N. D. (2017). *Mô hình hóa trong dạy học các tham số thống kê mô tả ở trường phổ thông [Modeling in teaching descriptive statistics parameters in high schools]*. Master thesis, Ho Chi Minh City University of Education.

- Phan, D. C. (Chief Editor), Ton, T. (Editor), Tran, D. C., Tran, P. D., & Tran, K. (2011). *Toan 7, tap 2 [Math 7 (Episode 2)]*. Hanoi: Vietnam Education Publishing House.
- Tran, V. H. (Chief Editor), Vu, T. (Editor), Doan, M. C., Do, M. H., & Nguyen, T. T. (2013). *Dai so 10 [Algebra 10]*. Hanoi: Vietnam Education Publishing House.
- Tran, V. H. (Chief Editor), Vu, T. (Editor), Doan, M. C., Do, M. H., & Nguyen, T. T. (2013). *Dai so 10, Sach giao vien [Algebra 10 (Teacher Textbook)]*. Hanoi: Vietnam Education Publishing House.

STATISTICS TEACHING: A COMPARISON STUDY BETWEEN HIGH SCHOOL PROGRAMME AND TEACHER TRAINING PROGRAM

*Nguyen Thi Nga**, *Tran Nguyen Khanh*

Ho Chi Minh City University of Education, Vietnam

**Corresponding author: Nguyen Thi Nga – Email: ngant@hcmue.edu.vn*

Received: April 28, 2020; Revised: May 07, 2020; Accepted: August 24, 2020

ABSTRACT

Statistics in 2018 Mathematics Curriculum has been significantly changed and will be taught from Grade 2 to Grade 12 instead of Grade 7 and Grade 10 as previously. This research aims to analyze the 2018 Mathematics Curriculum, in comparison with that of 2006 and the teacher education undergraduate program at the Department of Mathematics at Ho Chi Minh City University of Education. The results show that some contents are mentioned in the programme 2018 but have not appeared explicitly in the training students programme in 2016. A questionnaire was distributed to the students of the Department of Mathematics at the University to find out how they understand some new statistics knowledge mentioned in the 2018 Mathematics Curriculum and what difficulties they may face. Some solutions are suggested based on the difficulties shared by the students.

Keywords: Teaching statistics; 2018 Mathematics Curriculum; Descriptive Statistics