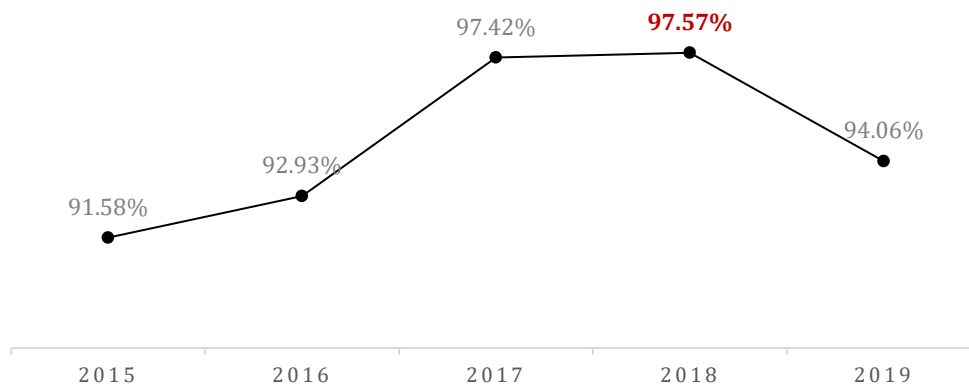


I. Đặt vấn đề

Kỳ thi trung học phổ thông quốc gia (THPTQG) là một kỳ thi quan trọng của học sinh trung học phổ thông ở Việt Nam được tổ chức vào năm 2015 – 2019, kỳ thi được tổ chức với 2 mục đích chính: Xét tốt nghiệp trung học phổ thông và tuyển sinh cao đẳng, đại học. Hằng năm, sau khi kỳ thi kết thúc thì Bộ GD&ĐT sẽ công bố phổ điểm môn thi và khối thi của toàn bộ thí sinh, đây cũng là công cụ giúp đánh giá chất lượng của kỳ thi và giúp thí sinh tự đánh giá năng lực của bản thân. Tuy nhiên, với tính chất của kỳ thi và quy định về xét tuyển đại học thì phổ điểm hiện nay chưa phản ánh đúng và đủ tin cậy do một số nguyên nhân sau:

- (1) **Tồn tại 2 nhóm thí sinh trong một môn thi:** Trong những năm tổ chức, 2018 là năm có nhiều biến động nhất khi là năm được đánh giá là đề thi có độ khó cao (6/9 môn thi có điểm trung bình dưới trung bình), nhưng tỉ lệ đỗ tốt nghiệp đạt 97.57%, cao nhất trong 5 năm tổ chức (nếu chỉ tính điểm thi tốt nghiệp bằng kỳ thi THPTQG thì chỉ có khoảng 48.6% thí sinh đạt). Điều này cho thấy để đạt điều kiện tốt nghiệp thì phần lớn kết quả xét tốt nghiệp phụ thuộc vào điểm học bạ khá lớn, tức là thí sinh không nhất thiết phải ôn thi đầy đủ 6 môn thi mà chỉ tập trung vào 3-4 môn thi dùng để xét tuyển cao đẳng, đại học thì cũng đạt 2 mục đích của kỳ thi. Do đó, một lượng lớn “thí sinh ảo” thi các môn không nằm trong nhóm môn chính của thí sinh đó sẽ góp phần làm lệch phổ điểm qua trái, tăng tỉ lệ thí sinh đạt điểm dưới trung bình và ảnh hưởng đến kết quả đánh giá độ khó của môn thi.



Biểu đồ 1. Tỉ lệ tốt nghiệp THPT năm 2015 – 2019

- (2) **Chưa xác định được số lượng thí sinh thực tế của mỗi khối thi:** Theo quy định, mỗi thí sinh chỉ trúng tuyển bởi một tổ hợp môn thi (Khối thi) cho một nguyện vọng duy nhất theo thứ tự ưu tiên và thí sinh có thể điều chỉnh lại tổ hợp, nguyện vọng sao cho phù hợp với kết quả sau khi kết thúc kỳ thi. Điều này đồng nghĩa với việc thí sinh sẽ chỉ lựa chọn tổ hợp môn thi có tổng điểm cao nhất để xét (Phần lớn các trường ĐH đều xét điểm một ngành với các tổ hợp là như nhau), và với việc tham gia nhiều môn thi như vậy thì mỗi thí sinh sẽ hình thành 3,4 khối thi. Thực tế hiện nay, phổ điểm các tổ hợp xét tuyển đại học đang được tính dựa trên toàn bộ thí sinh có hình thành điểm thi của khối thi, điều này làm cho việc xác định số lượng thí sinh thực tế và so sánh chất lượng giữa các khối với nhau là không thể thực hiện.

Từ những lý do trên, bài nghiên cứu này sẽ tập trung vào việc:

- ✓ **Điều chỉnh dữ liệu điểm thi thông qua lọc “thí sinh ảo”.**
- ✓ **So sánh và phân tích dữ liệu điểm thi mới để đánh giá chất lượng của kỳ thi.**

Dựa trên bộ dữ liệu “Điểm thi THPT Quốc Gia 2018” và các công cụ hỗ trợ (Python, SQL, MySQL, Power BI, Excel).

II. Khái quát về dữ liệu

1. Nguồn dữ liệu

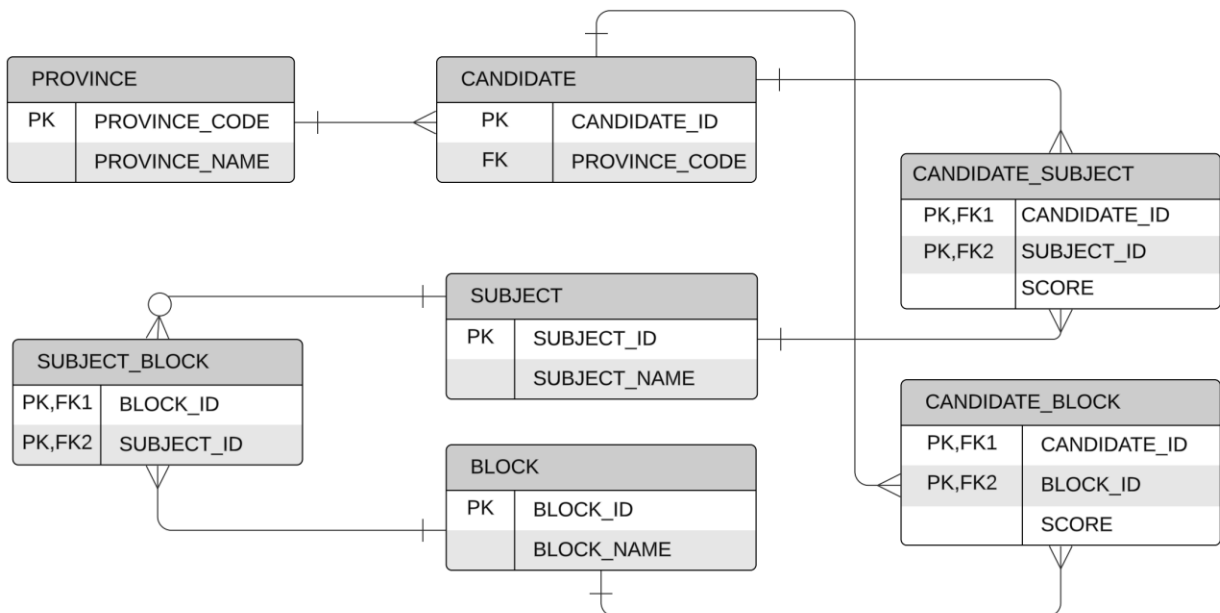
- Trích xuất dữ liệu điểm thi từ API của trang tra cứu điểm thi TTXVN với đầu vào là số báo danh và năm thi.

```
<script type="application/ld+json">
{
  "@context": "https://schema.org/",
  "@type": "WebSite",
  "name": "Điểm chuẩn đại học - Điểm thi THPT quốc gia 2018 - Chỉ tiêu các trường ĐH-CĐ năm 2018",
  "url": "https://diemthi.vnanet.vn",
  "potentialAction": {
    "@type": "SearchAction",
    "target": "https://diemthi.vnanet.vn/Home/SearchBySobaodanh?code={search_term_string}&nam=2018",
    "query-input": "required name=search_term_string"
  }
}
```

2. Quá trình thu thập dữ liệu



3. Mô hình hóa dữ liệu



Biểu đồ 2. Mô hình quan hệ thực thể (ERD Diagram) Crow's foot notation cho cơ sở dữ liệu điểm thi THPTQG

III. Phương pháp nghiên cứu

1. Thống kê mô tả

1.1. Các tham số thống kê đo lường độ tập trung (Measures of Central Tendency)

- a) **Giá trị trung bình (Mean):** Là giá trị trung bình số học của một biến, được tính bằng tổng các giá trị quan sát chia cho số quan sát. Giá trị trung bình được tính bằng công thức sau:

$$\mu = \frac{\sum x_i}{N}$$

- b) **Yếu vị (Mode):** Là giá trị có tần suất xuất hiện lớn nhất của một tập hợp các số đo, dạng này thường được dùng đối với dạng dữ liệu thang biểu danh.

- c) **Trung vị (Median):** Là số nằm giữa (nếu lượng quan sát là số lẻ) hoặc là giá trị trung bình của hai quan sát nằm giữa (nếu số lượng quan sát là số chẵn) của một dãy quan sát được sắp xếp theo thứ tự từ nhỏ đến lớn.

$$Med(x) = \begin{cases} x_{\left[\frac{n+1}{2}\right]} & \text{nếu } n \text{ là số chẵn} \\ \frac{x_{\left[\frac{n}{2}\right]} + x_{\left[\frac{n}{2} + 1\right]}}{2} & \text{nếu } n \text{ là số lẻ} \end{cases}$$

1.2. Các tham số thống kê đo lường mức độ phân tán (Measures of Dispersion)

- d) **Phương sai (Variance):** Dùng để đo lường mức độ phân tán của một tập các giá trị quan sát xung quanh giá trị trung bình của tập quan sát đó. Phương sai bằng trung bình các bình phương sai lệch giữa các giá trị quan sát đối với giá trị trung bình của các quan sát đó.

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

- e) **Độ lệch chuẩn (Standard Deviation):** Độ lệch chuẩn chính bằng căn bậc hai của phương sai. Vì phương sai là trung bình của các bình phương sai lệch của các giá trị quan sát từ giá trị trung bình, việc khảo sát phương sai thường cho các giá trị rất lớn, do đó sử dụng phương sai sẽ gặp khó khăn trong việc diễn giải kết quả.

$$\sigma = \sqrt{\sigma^2}$$

- f) **Hệ số đối xứng (Skewness):** Cho biết dạng phân phối của các giá trị quan sát có thể được sử dụng để kiểm nghiệm tính phân phối chuẩn. Một giá trị Skewness dương cho biết nhánh của phân phối này lệch qua bên phải và ngược lại một trị âm chỉ ra nhánh của phân phối này dài qua bên trái.

$$Skewness = \frac{n}{(n-1)(n-2)} \sum \left(\frac{x_i - \bar{x}}{\sigma} \right)^3$$

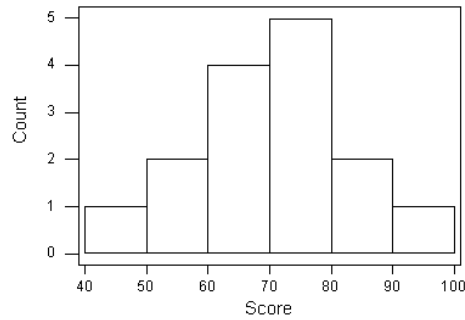
- g) **Hệ số tập trung (Kurtosis):** Dùng để đánh giá đỉnh của đường cong quan sát với dạng đường cong phân phối chuẩn. Nếu Kurtosis > 0 thì cho thấy xu hướng tập trung các quan sát xung quanh giá trị trung bình, phân phối có đỉnh, ngược lại thì đường cong có dạng hẹp hơn hay đường cong của phân phối tương đối bằng phẳng.

$$Kurtosis = \left\{ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{x_i - \bar{x}}{\sigma} \right)^4 \right\} - \frac{3(n-1)^2}{(n-3)(n-5)}$$

2. Trực quan hóa dữ liệu

2.1. Biểu đồ Histogram

- Histogram (Biểu đồ tần suất) là biểu đồ với một đại diện gần đúng của phân phối dữ liệu số. Biểu đồ này được giới thiệu lần đầu bởi Karl Pearson.
- Biểu đồ tần suất gồm 2 phần:
 - + Trục x: Chia khoảng (binning) các giá trị. Các khoảng này phải đều nhau và thường có kích cỡ như nhau.
 - + Trục y: Đếm tần số xuất hiện của giá trị trong dữ liệu ứng với mỗi khoảng.



Biểu đồ 3. Mô tả biểu đồ Histogram

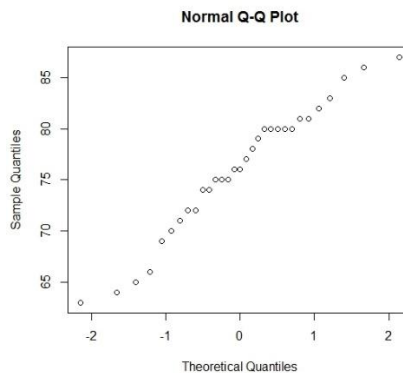
- Xác định các bin cho các biểu đồ môn thi và khối thi, dựa vào tính chất của bài thi (trắc nghiệm, tự luận), các bins được xác định như sau:

	Hình thức thi	Điểm mỗi câu	Độ lớn mỗi khoảng
Toán	Trắc nghiệm (50)	0.2	0.2
Ngữ Văn	Tự luận	-	0.25
Tiếng Anh	Trắc nghiệm (50)	0.2	0.2
Vật Lý	Trắc nghiệm (40)	0.25	0.25
Hóa Học	Trắc nghiệm (40)	0.25	0.25
Sinh Học	Trắc nghiệm (40)	0.25	0.25
Lịch Sử	Trắc nghiệm (40)	0.25	0.25
Địa Lý	Trắc nghiệm (40)	0.25	0.25
Giáo dục Công dân	Trắc nghiệm (40)	0.25	0.25
Khối thi (A00, A01, B00, C00, D01)	-	-	0.5

Bảng 1. Xác định khoảng giá trị (bins) cho dữ liệu điểm thi

2.2. Biểu đồ QQ-Plot (Quantile–Quantile Plot)

- Biểu đồ QQ-Plot là biểu đồ giúp đánh giá xem một tập dữ liệu có tuân theo phân phối lý thuyết như phân phối chuẩn hay không.
- Trục hoành của biểu đồ là định lượng theo luật phân phối chuẩn (theoretical quantile) và trục hoành định lượng của số liệu (sample quantiles). Nếu phân phối của biến theo luật phân phối chuẩn, thì đường biểu diễn phải theo một đường thẳng chéo 45 độ (tức là định lượng phân phối và định lượng số liệu bằng nhau).



Biểu đồ 4. Mô tả biểu đồ Normal Q-Q Plot

IV. Kết quả nghiên cứu và thảo luận

1. Lọc dữ liệu điểm thi

- Giả định khối thi có điểm thi cao nhất của mỗi thí sinh là khối thi thí sinh lựa chọn và 3 môn thi của khối thi đó là 3 môn thí sinh lựa chọn để thi. Trường hợp có 2 khối thi bằng điểm nhau thì lấy tất cả các môn đó.

- Với hơn 90% thí sinh lựa chọn các khối thi truyền thống A00, A01, B00, C00 và D01 nên sẽ lựa chọn các khối thi này để xét.

Ví dụ: Thí sinh **01000001** có điểm thi như sau:

Candidate_id	Subject_id	Score
01000001	Toan	8.0
01000001	Anh	7.4
01000001	Van	6.5
01000001	Ly	6.5
01000001	Hoa	3.5
01000001	Sinh	3.5

Candidate_id	Block_id	Score
01000001	A00	18.0
01000001	A01	21.9
01000001	B00	15.0
01000001	D01	21.9

Bảng 2. Cấu trúc điểm thi và khối thi lưu trữ trong cơ sở dữ liệu

Trong trường hợp này khối **A01** và **D01** của thí sinh sẽ được giữ lại đồng thời giữ lại các môn **Toán, Ngữ Văn, Tiếng Anh và Vật Lý** để xét. Các khối thi và điểm thi còn lại sẽ được xem là các giá trị ngoại lai (Outliers).

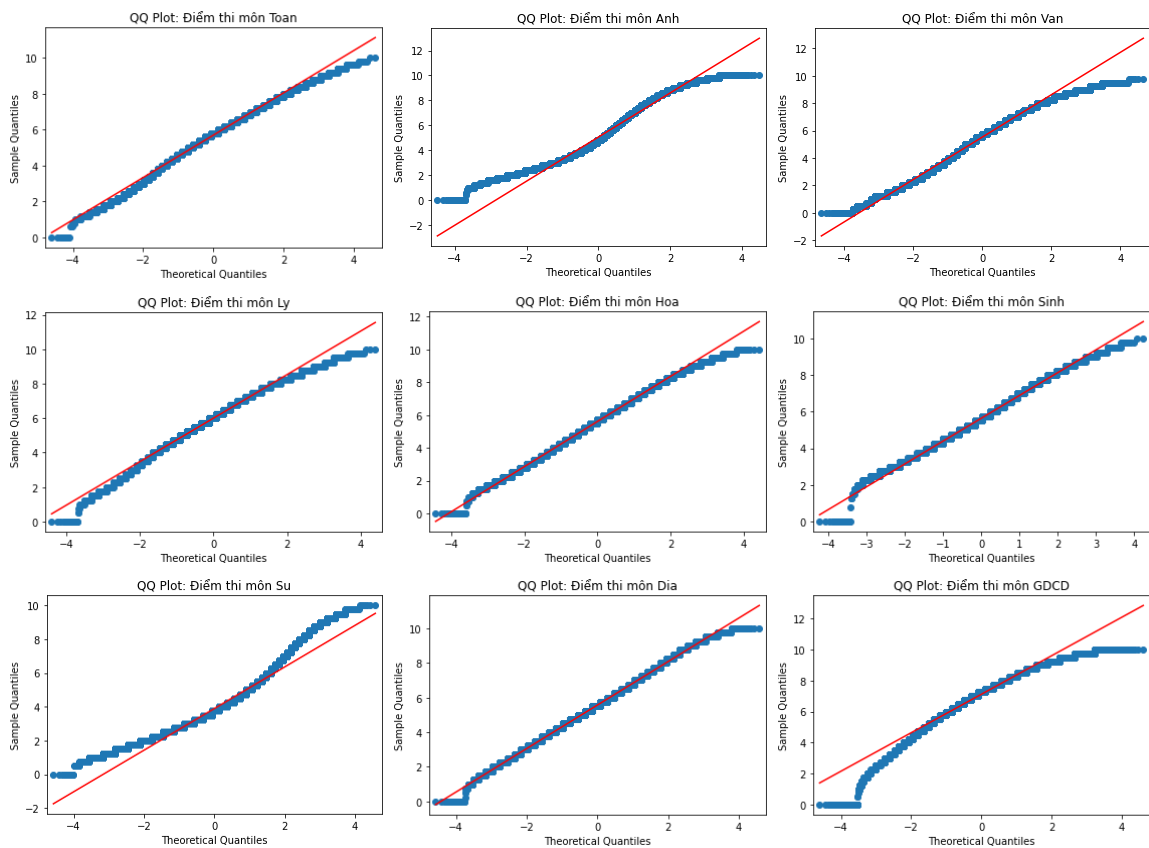
2. Điểm thi các môn thi

	Toán	Ngữ Văn	Tiếng Anh	Vật Lý	Hóa Học	Sinh Học	Lịch Sử	Địa Lí	GD&CD
Mean	4.88	5.46	3.92	4.97	4.86	4.54	3.76	5.46	7.12
Median	5.00	5.50	3.60	5.00	4.75	4.50	3.50	5.50	7.25
Mode	5.40	6.00	3.00	5.00	5.00	4.25	3.25	5.25	7.50
Standard Deviation	1.46	1.46	1.58	1.55	1.54	1.25	1.21	1.26	1.24
Sample Variance	2.14	2.15	2.49	2.40	2.37	1.56	1.47	1.60	1.55
Kurtosis	-0.50	-0.28	0.92	-0.49	-0.40	0.35	1.29	0.16	0.65
Skewness	-0.08	-0.23	1.07	0.03	0.09	0.35	0.87	0.03	-0.62
Range	10.00	9.75	10.00	10.00	10.00	10.00	10.00	10.00	10.00
Minimum	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Maximum	10.00	9.75	10.00	10.00	10.00	10.00	10.00	10.00	10.00
Count	919139	903659	834690	403167	405860	399728	553056	552492	482763

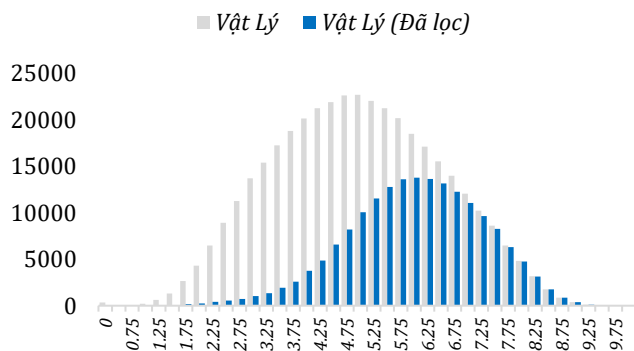
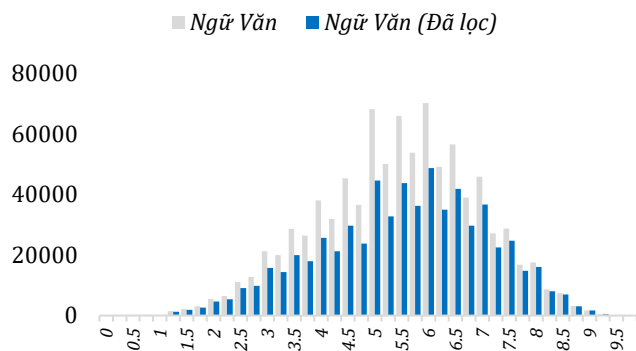
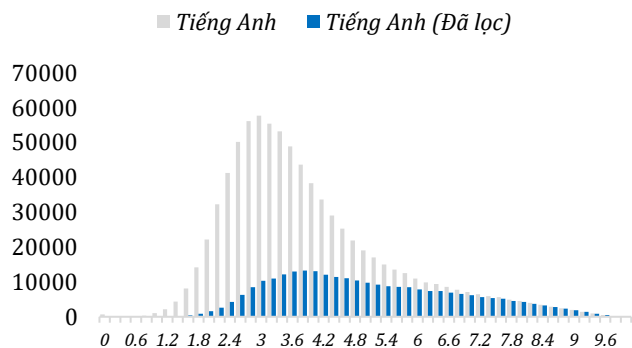
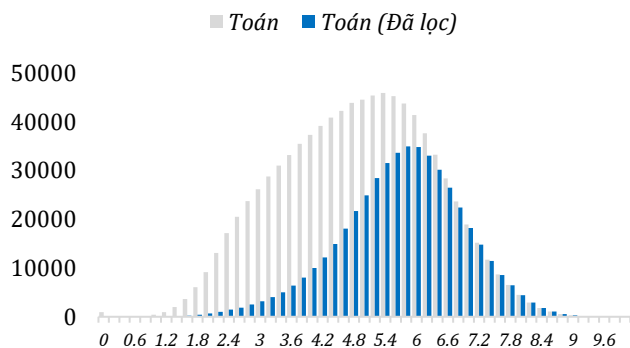
Bảng 3. Mô tả thống kê điểm thi của các môn thi (trước khi lọc)

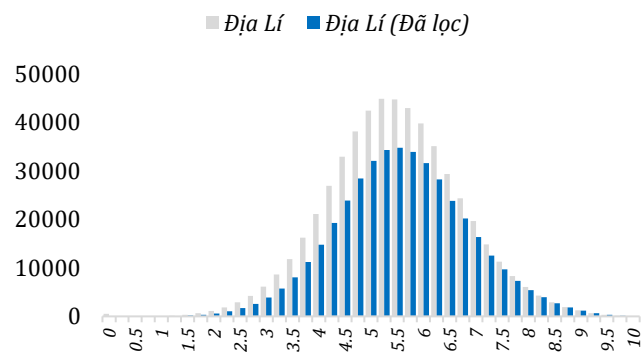
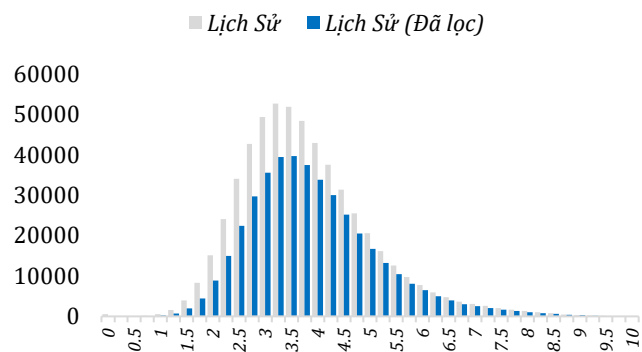
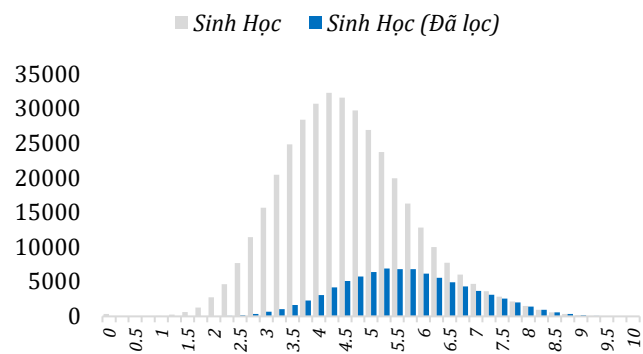
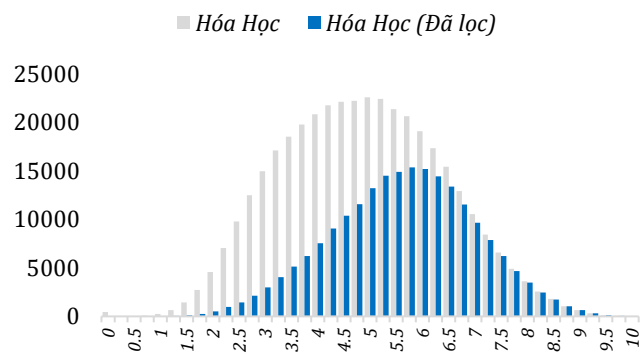
	Toán	Ngữ Văn	Tiếng Anh	Vật Lý	Hóa Học	Sinh Học	Lịch Sử	Địa Lí	GD&CD
Mean	5.70	5.53	5.06	6.00	5.61	5.64	<u>3.89</u>	5.57	7.12
Median	5.80	5.75	4.80	6.00	5.75	5.50	<u>3.75</u>	5.50	7.25
Mode	5.80	6.00	3.80	6.00	5.75	5.25	<u>3.50</u>	5.50	7.50
Standard Deviation	1.18	1.55	1.77	1.26	1.38	1.25	1.23	1.26	1.24
Sample Variance	1.39	2.39	3.13	1.60	1.89	1.55	1.52	1.58	1.55
Kurtosis	0.23	<u>-0.38</u>	<u>-0.63</u>	0.07	<u>-0.20</u>	<u>-0.19</u>	1.24	0.10	0.65
Skewness	<u>-0.31</u>	<u>-0.28</u>	0.45	<u>-0.30</u>	<u>-0.11</u>	0.11	0.90	0.08	<u>-0.62</u>
Range	10.00	9.75	10.00	10.00	10.00	10.00	10.00	10.00	10.00
Minimum	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Maximum	10.00	9.75	10.00	10.00	10.00	10.00	10.00	10.00	10.00
Count	483192	651258	272274	179777	213564	87421	423252	423252	482763

Bảng 4. Mô tả thống kê điểm thi của các môn thi (sau khi lọc)

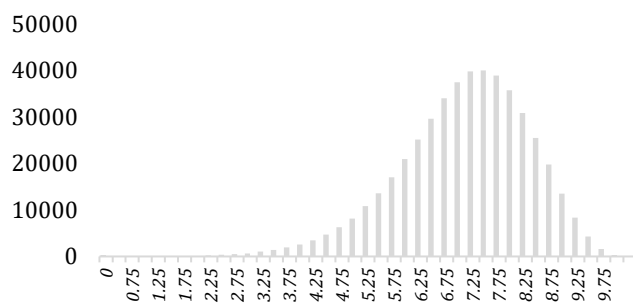


Biểu đồ 5. Biểu đồ QQ-Plot điểm thi các môn thi

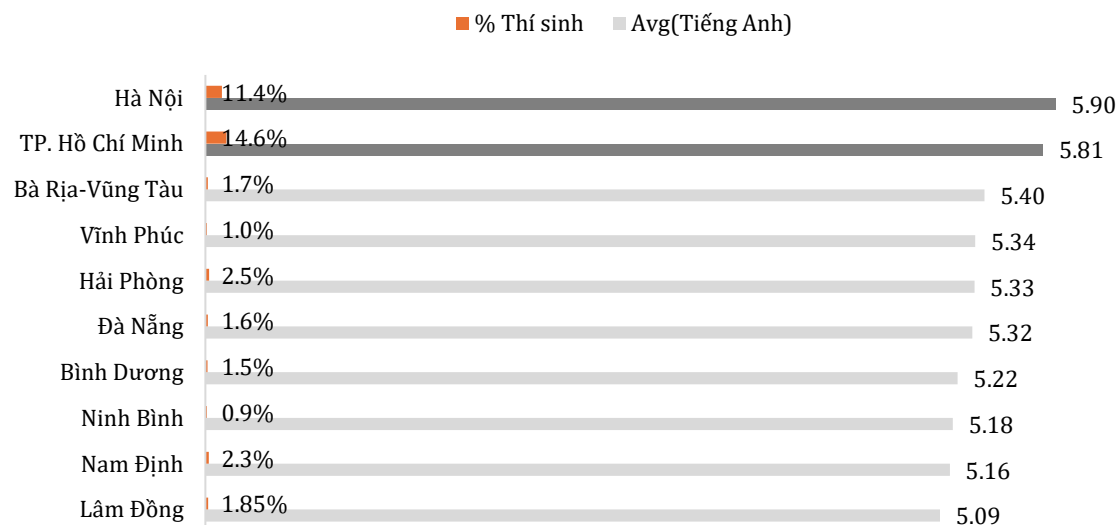




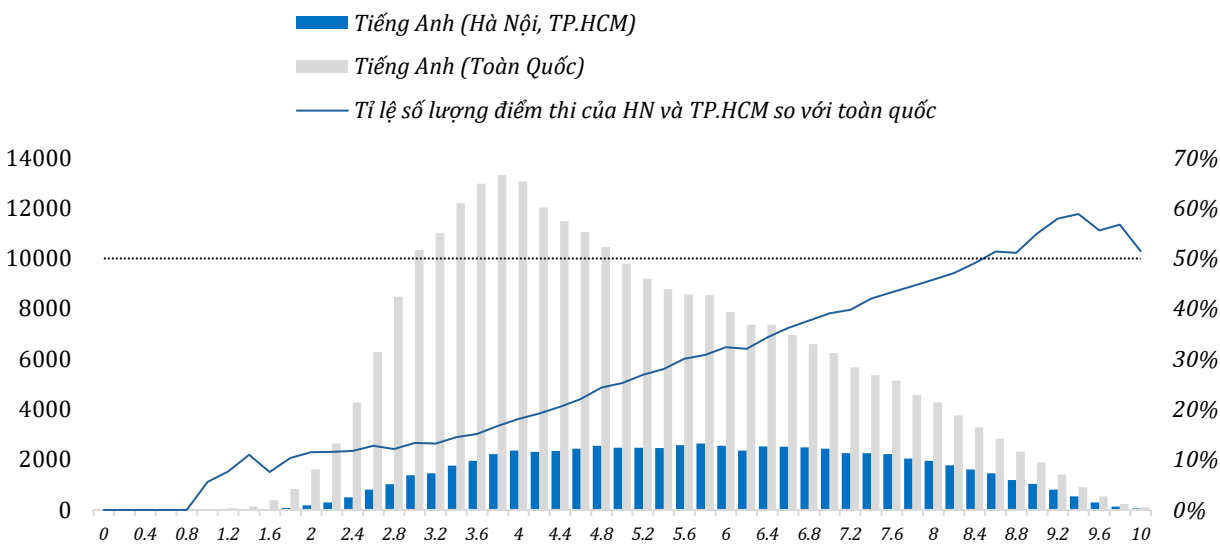
GD CD



Biểu đồ 6. Biểu đồ Histogram cho các môn thi



Biểu đồ 7. Biểu đồ thể hiện điểm trung bình môn tiếng Anh của 10 tỉnh/thành có điểm cao nhất



Biểu đồ 8. Biểu đồ so sánh phổ điểm môn tiếng Anh của thí sinh toàn quốc và Hà Nội, TP.HCM

Nhận xét: Dữ liệu điểm thi sau khi lọc

- **Về độ khó của đề thi:** Dựa vào chỉ số Mean, Median, Mode, STD kết hợp với biểu đồ Histogram. Phần lớn điểm các môn thi đều nằm ở mức 5-6 điểm và phân phối có đỉnh, riêng môn Tiếng Anh và Lịch Sử có sự khác biệt, cụ thể:
 - + **Môn tiếng Anh:** Điểm trung bình là 5.06 (± 1.77) nhưng có đỉnh ở mức điểm 3.8 cho thấy có sự phân tán lớn ở phía mức điểm cao hơn, đỉnh phổ điểm không hiện diện rõ ràng và khó để xác định đề thi khó hay dễ.
 - + **Môn Lịch sử:** Với điểm trung bình là 3.89 (± 1.23) (thấp nhất trong các môn thi) và đỉnh phổ điểm ở mức 3.5 cho thấy đề thi môn Lịch sử ở mức khó, gây bất lợi với nhóm thí sinh chọn môn Lịch sử để xét tuyển cao đẳng, đại học.
 - + **Môn GD&ĐT:** Với điểm trung bình là 7.12 (± 1.24) (cao nhất trong các môn thi) với đỉnh phổ điểm ở 7.25 cho thấy đề thi môn GD&ĐT tương đối dễ và là môn không (ít) nằm trong các khối thi xét tuyển cao đẳng đại học nên đây là mức điểm phù hợp để xét tốt nghiệp.
- **Về phân phối của phổ điểm:** Dựa vào chỉ số Kurtosis, Skewness kết hợp với biểu đồ QQ-Plot
 - + Các môn **Toán, Ngữ Văn, Vật Lý, Hóa Học, Sinh Học, Địa Lý** có phân phối gần với phân phối chuẩn nhất (có giá trị Kurtosis và Skewness tiệm cận với 0), trong đó môn Ngữ Văn là môn tự luận nên có đỉnh thấp hơn so với phân phối chuẩn.
 - + Phổ điểm môn **tiếng Anh** có đỉnh thấp và dữ liệu tập trung phía trái nhưng không quá nhiều, các mức điểm đổ dốc từ trung vị về phía phải của phổ điểm.
 - + Phổ điểm môn **Lịch sử** có đỉnh cao và dữ liệu tập trung về phía trái và tạo thành hình đuôi dài về phía phải.
- **Phân tích về điểm thi môn tiếng Anh:** Xét về khu vực thì Hà Nội và Thành phố Hồ Chí Minh là 2 tỉnh thành có điểm thi môn tiếng Anh vượt trội hơn so với các tỉnh/thành khác với lần lượt điểm trung bình là 5.9 và 5.81, cao hơn trung bình toàn quốc là 5.06. Về số lượng thí sinh, tổng thí sinh dự thi ở Hà Nội và Tp. Hồ Chí Minh chiếm 26% (1/4 so với tổng thí sinh). Ở biểu đồ 8 mô tả số lượng và tỉ lệ bài thi của 2 tỉnh thành so với toàn quốc ở mỗi mức điểm. Xét ở ngưỡng 50%, số lượng bài thi đạt điểm cao hơn so với toàn quốc nằm ở mức điểm 8.6 – 10 và phổ điểm có xu hướng phẳng hơn so với phổ điểm toàn quốc.

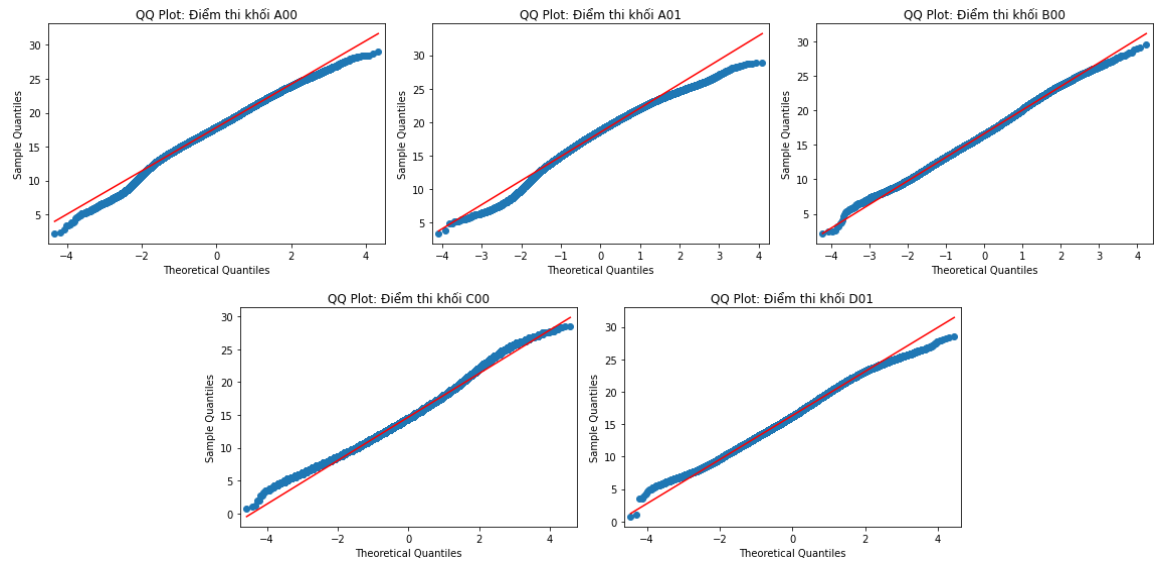
3. Điểm thi các khối thi

	A00	A01	B00	C00	D01
Mean	15.56	15.05	15.14	14.59	14.48
Median	15.65	14.90	15.15	14.50	14.35
Mode	16.05	14.80	15.10	14.25	14.20
Standard Deviation	3.70	3.49	3.24	3.28	3.41
Sample Variance	13.66	12.21	10.48	10.74	11.60
Kurtosis	-0.11	-0.05	0.30	0.13	-0.15
Skewness	-0.11	0.14	0.00	0.26	0.23
Range	29.05	28.95	29.55	28.50	28.60
Minimum	0.00	0.00	0.00	0.00	0.00
Maximum	29.05	28.95	29.55	28.50	28.60
Count	383008	368666	382144	548558	811965

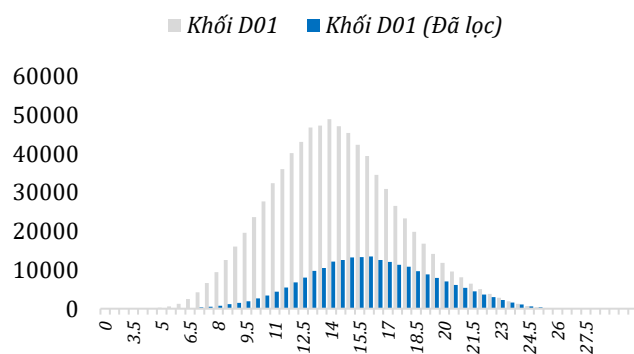
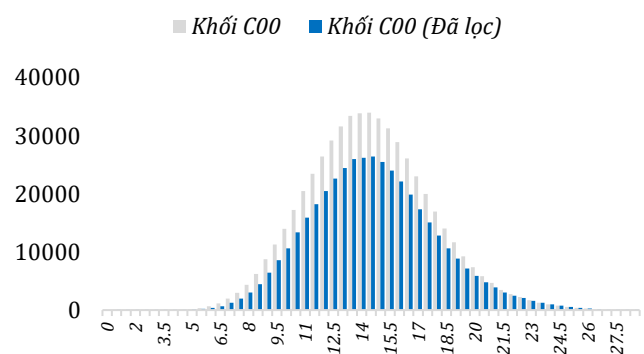
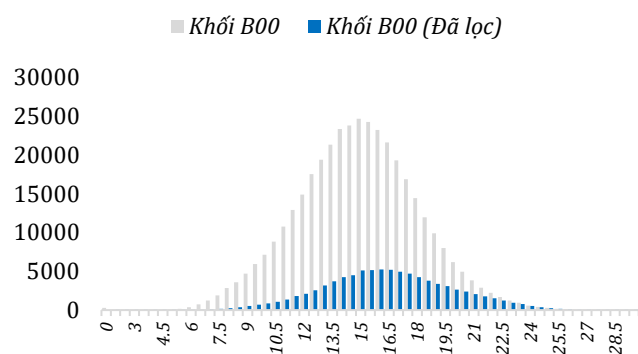
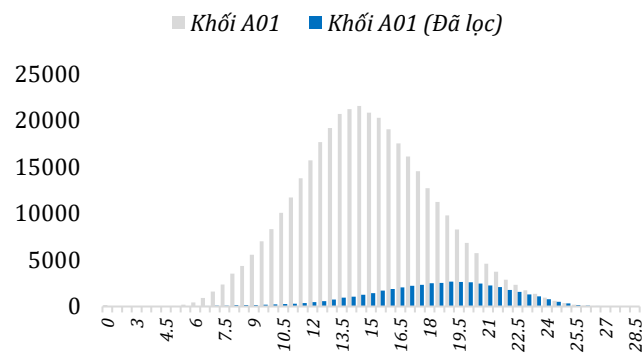
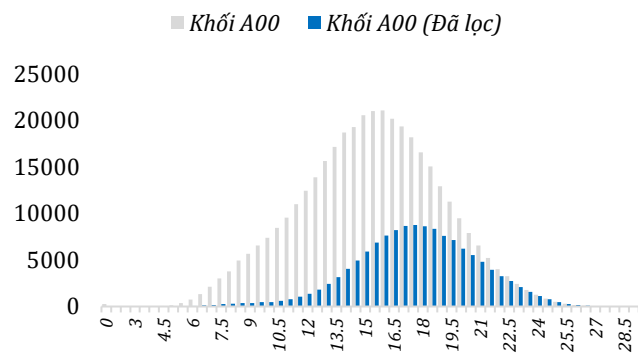
Bảng 5. Mô tả thống kê điểm khối thi (toàn bộ thí sinh)

	A00	A01	B00	C00	D01
Mean	17.82	18.50	16.65	14.67	16.34
Median	17.85	18.80	16.55	14.50	16.20
Mode	17.70	19.15	16.55	14.25	15.60
Standard Deviation	3.20	3.60	3.42	3.31	3.40
Sample Variance	10.23	12.97	11.71	10.96	11.55
Kurtosis	0.44	0.29	-0.13	0.18	-0.29
Skewness	-0.27	-0.52	0.09	0.34	0.05
Range	26.85	25.55	27.35	27.75	27.85
Minimum	2.20	3.40	2.20	0.75	0.75
Maximum	29.05	28.95	29.55	28.50	28.60
Count	133947	46580	87421	423252	230956

Bảng 6. Mô tả thống kê điểm thi khối thi (sau khi lọc)



Biểu đồ 9. Biểu đồ QQ-Plot của các khối thi

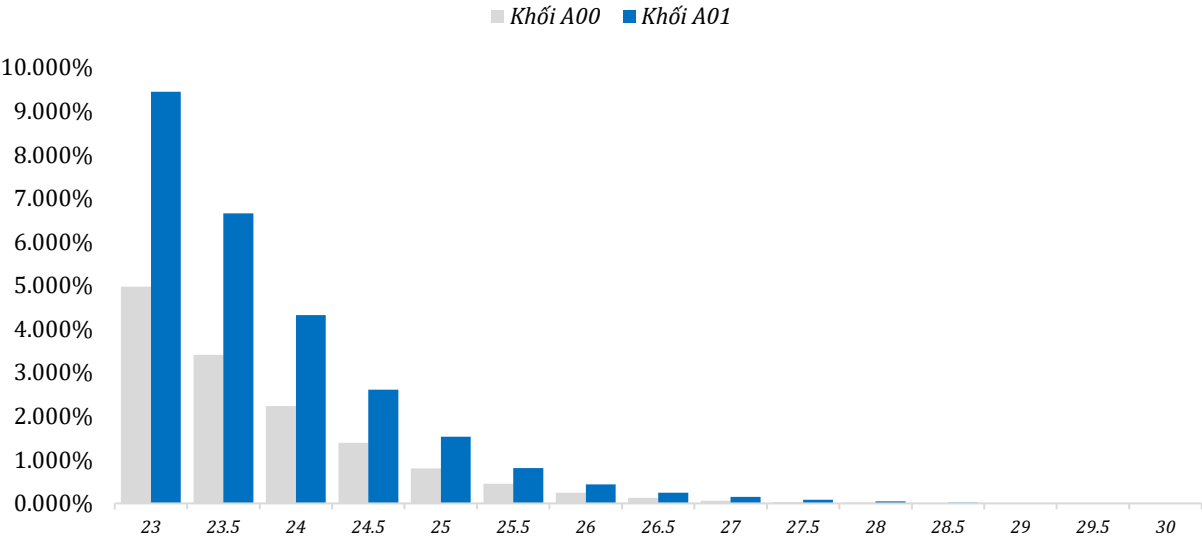


Biểu đồ 10. Biểu đồ Histogram cho khối thi

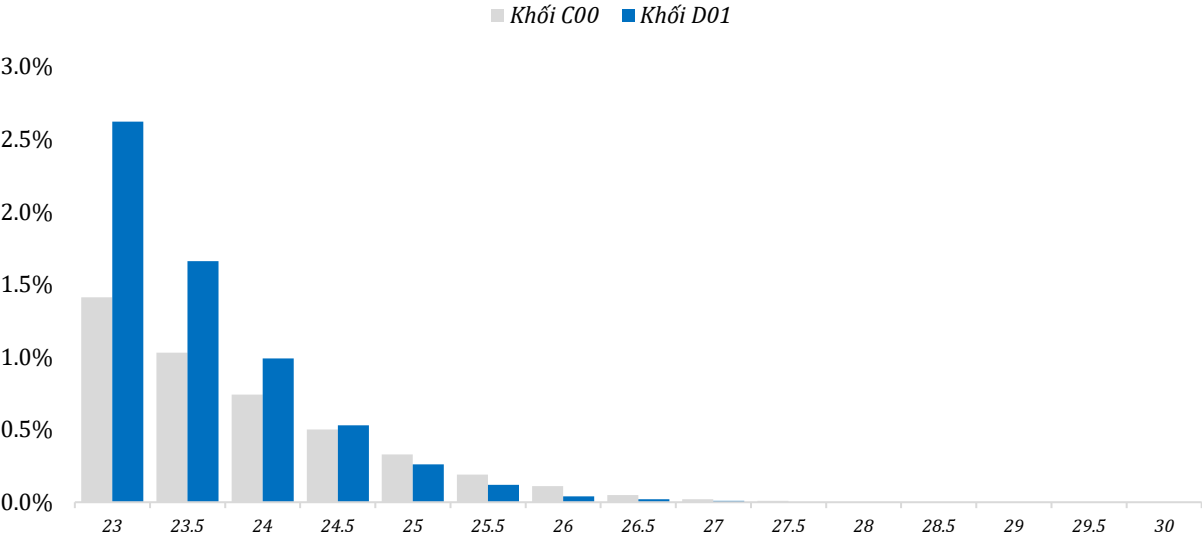
Mức điểm	A00		A01		B00		C00		D01	
	Số lượng	%	Số lượng	%	Số lượng	%	Số lượng	%	Số lượng	%
30	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
29.5	0	0.00%	0	0.00%	1	0.00%	0	0.00%	0	0.00%
29	1	0.00%	0	0.00%	3	0.00%	0	0.00%	0	0.00%
28.5	2	0.00%	10	0.02%	4	0.00%	2	0.00%	1	0.00%
28	20	0.01%	23	0.05%	12	0.01%	7	0.00%	4	0.00%
27.5	45	0.03%	40	0.09%	25	0.03%	31	0.01%	9	0.00%
27	85	0.06%	71	0.15%	50	0.06%	84	0.02%	17	0.01%
26.5	170	0.13%	116	0.25%	100	0.11%	201	0.05%	43	0.02%
26	332	0.25%	204	0.44%	199	0.23%	455	0.11%	103	0.04%
25.5	603	0.45%	378	0.81%	359	0.41%	824	0.19%	281	0.12%
25	1074	0.80%	714	1.53%	622	0.71%	1383	0.33%	597	0.26%
24.5	1859	1.39%	1217	2.61%	992	1.13%	2136	0.50%	1213	0.53%
24	2993	2.23%	2013	4.32%	1509	1.73%	3116	0.74%	2276	0.99%
23.5	4571	3.41%	3101	6.66%	2286	2.61%	4372	1.03%	3835	1.66%

23	6659	4.97%	4399	9.44%	3246	3.71%	5987	1.41%	6062	2.62%
22.5	9400	7.02%	5982	12.84%	4477	5.12%	8083	1.91%	9012	3.90%
22	12666	9.46%	7750	16.64%	5984	6.85%	10557	2.49%	12672	5.49%
21.5	16622	12.41%	9834	21.11%	7753	8.87%	13600	3.21%	17110	7.41%
21	21444	16.01%	12102	25.98%	9837	11.25%	17497	4.13%	22513	9.75%
20.5	26982	20.14%	14569	31.28%	12221	13.98%	22274	5.26%	28649	12.40%
20	33198	24.78%	17174	36.87%	14877	17.02%	28148	6.65%	35666	15.44%
19.5	40355	30.13%	19823	42.56%	17976	20.56%	35303	8.34%	43623	18.89%
19	47956	35.80%	22498	48.30%	21379	24.46%	44180	10.44%	52475	22.72%
18.5	56301	42.03%	25041	53.76%	25197	28.82%	54787	12.94%	62141	26.91%
18	64928	48.47%	27543	59.13%	29456	33.69%	67616	15.98%	72963	31.59%
17.5	73696	55.02%	29884	64.16%	34152	39.07%	82683	19.54%	84260	36.48%
17	82358	61.49%	32112	68.94%	39122	44.75%	100052	23.64%	96363	41.72%
16.5	90578	67.62%	34173	73.36%	44329	50.71%	119957	28.34%	108912	47.16%
16	98226	73.33%	36063	77.42%	49574	56.71%	142074	33.57%	122383	52.99%
15.5	105115	78.48%	37787	81.12%	54722	62.60%	166096	39.24%	135663	58.74%
15	111019	82.88%	39240	84.24%	59856	68.47%	191593	45.27%	148896	64.47%

Bảng 7. Bảng thống kê số lượng và tỉ lệ thí sinh đạt trên mức điểm tương ứng



Biểu đồ 11. So sánh phổ điểm khối A00 và A01 ở mức điểm 23 trở lên theo tỉ lệ



Nhận xét: Dữ liệu điểm thi sau khi lọc

- **Về dạng phổ điểm:** Phổ điểm của các khối thi gần với phân phối chuẩn, có các giá trị mean, median và mode gần bằng nhau. Tuy nhiên chênh lệch điểm giữa phổ điểm là khá lớn, trong đó điểm khối **A01** cao nhất với 18.5 (± 3.6) và thấp nhất là khối **C00** với 14.67 (± 3.31).
- **So sánh khối A00 và khối A01:** Tuy hình thức chọn tổ hợp (khối thi) khác nhau nhưng về cơ bản phương pháp xét tuyển là như nhau. Khối A00 và khối A01 là hai khối phổ biến xét tuyển vào các khối ngành kỹ thuật, công nghệ. Dựa theo bảng 7, xét từ khoảng 15 điểm (mức điểm phù hợp để xét tuyển vào cao đẳng, đại học), số lượng thí sinh khối A00 là 111,019 và khối A01 là 39,240 (khối A00 chiếm gần $\frac{3}{4}$ so với tổng thí sinh 2 khối). Mặc dù số lượng thí sinh chọn khối A01 thấp nhưng điểm thi ở các mức điểm càng cao thì khối A01 càng có lợi hơn, đặc biệt là ở mức điểm 23 trở lên (mức điểm dùng để xét tuyển các ngành thường có điểm chuẩn cao), tỉ lệ thí sinh đạt các mức điểm này cao gấp 2 lần tính theo số lượng thí sinh mỗi khối thi.
- **So sánh khối C00 và khối D01:** Xét từ khoảng 15 điểm, số lượng thí sinh chọn khối C00 là 191,593, khối D01 là 148,896, 2 khối có số lượng thí sinh gần tương đương nhau. Mặc dù với điểm trung bình thấp nhưng tỉ lệ số lượng thí sinh khối C00 cao hơn khối D01 từ mức điểm 24 trở đi, trong khi đó khối D01 chiếm ưu thế hơn ở mức điểm 15 – 23.

V. Kết luận

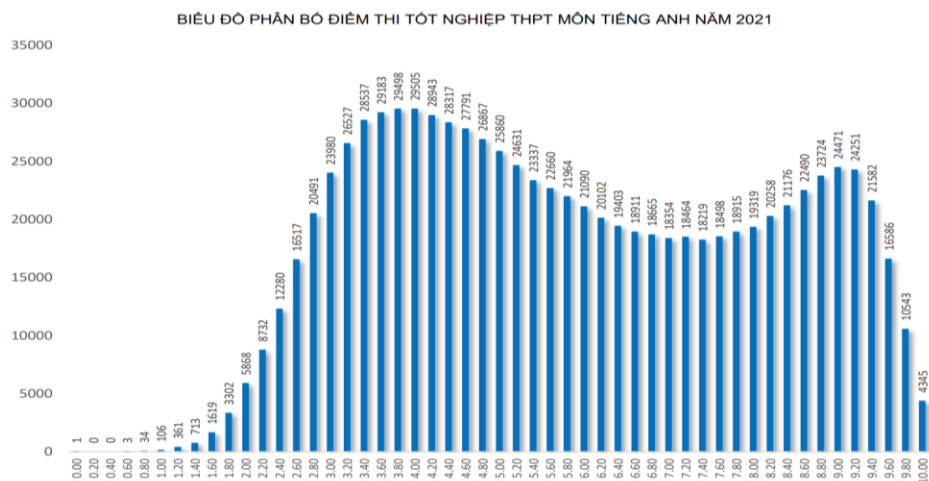
Đề thi THPT Quốc Gia 2018 chỉ phù hợp cho xét tuyển CĐ, ĐH: Vì mục đích của kỳ thi là vừa xét tốt nghiệp và vừa xét tuyển CĐ, ĐH nên việc điều chỉnh đề thi thỏa mãn cả 2 yêu cầu là rất khó, khi phải vừa đảm mức thí sinh đạt điểm trên trung bình cao và phải hạn chế được điểm 10. Mức điểm 6.0 - 10.0 chưa đủ rộng để kiểm soát độ khó của đề thi. Nếu đề dễ sẽ tăng tỷ lệ thí sinh đạt trung bình cao thì sẽ vô tình tịnh tiến phổ điểm sang phải làm số lượng điểm giỏi tăng cao.

Cần giảm số lượng môn thi THPT Quốc Gia: Như đã đề cập ở phần đầu, mặc dù kỳ thi tổ chức để tránh học sinh học tủ, học lệch nhưng hiện tượng này vẫn diễn ra và không tránh khỏi. Một phần vì điểm học bạ cao có thể “bù” cho các môn thí sinh không chọn là môn xét tuyển Đại Học và để tập trung cho mục tiêu chính là ôn thi 3 môn để xét tuyển Đại Học, đặc biệt là nhóm thí sinh chọn tổ hợp Khoa học Tự Nhiên (thí sinh khối A01 chỉ học môn Lý hoặc thí sinh khối D01 không học môn thi nào). Dựa vào bảng 8, có thể thấy số môn thi càng ít thì điểm chênh lệch thấp hơn.

Số lượng môn thi	Tổng số thí sinh	Tỉ lệ thí sinh có mức chênh lệch điểm thi các môn (Môn điểm thi cao nhất – Môn điểm thi thấp nhất) [Không bao gồm môn GD&CD]		
		≤ 1	≤ 2	≤ 3
1	480	100%	100%	100%
2	955	71%	91%	97%
3	37739	47%	79%	94%
4	63443	19%	56%	85%
5	7657	15%	46%	74%
6	769823	4%	25%	61%
7	5759	6%	35%	72%
8	296	4%	15%	51%
9	35756	1%	15%	48%

Bảng 8. Thống kê về mức điểm chênh lệch điểm thi của thí sinh (dữ liệu trước khi lọc)

Cần đảm bảo độ khó đồng đều giữa các môn thi: Qua việc phân tích trên, môn tiếng Anh là môn khó có thể điều chỉnh nhất vì chưa xác định được đỉnh phổ điểm. Việc điều chỉnh đề thi dễ hơn có thể dẫn đến tình trạng mưa điểm 10 như năm 2017 hay tệ hơn là hình thành một phổ điểm có 2 đỉnh ở nhóm điểm thấp và nhóm điểm cao như năm 2021. Đặc biệt đây là môn thi có sự đầu tư khác nhau giữa các khu vực nông thôn và thành thị.



Biểu đồ 13. Phổ điểm môn tiếng Anh 2021

VI. Tài liệu tham khảo

Bộ GD&ĐT. (2017). **Thông tư “Ban hành Quy chế thi THPT Quốc Gia và xét công nhận thi tốt nghiệp THPT”**.
https://moet.gov.vn/content/vanban/lists/vbpq/attachments/1314/02_vbhn_bgddt.pdf

Huy N. D. & Cáp, D.T. (2012). **Giáo trình Xác suất và Thống kê**. NXB ĐHQG TP HCM.
<https://ir.vnulib.edu.vn/handle/VNUHCM/5478>

Trần, D. P., & Vũ, C. T. (2023). **Tiêu chuẩn định lượng cho một bảng điểm kết thúc học phần đẹp**. *Tạp Chí Nghiên cứu Khoa học Và Phát triển*, 2(1), 35-40.
<https://doi.org/10.58902/tcnckhpt.v2i1.33>

Hiep, P. (2015, Aug 18). **Thế nào là một phổ điểm đẹp?**. *Tạp Chí Tia Sáng*. Retrieved Aug 23, 2023, from <https://tiasang.com.vn/giaoduc/the-nao-la-mot-pho-diem-dep-8947/>

Anh, H. (2018, Jul 16). **Tỷ lệ đỗ tốt nghiệp THPT quốc gia 2018 cao nhất trong 4 năm gần đây**. *Tạp chí Kinh tế và Dự báo*. Retrieved Aug 23, 2023, from <https://kinhtevadubao.vn/ty-le-do-tot-nghiepthpt-quoc-gia-2018-cao-nhat-trong-4-nam-gan-day-13027.htm>