




# PHÂN TÍCH ĐIỂM THI THPT QUỐC GIA NĂM 2018

Thực hiện: Huỳnh Trọng Khoa

 [Github.com/khoaht312/vnhsge-2018](https://github.com/khoaht312/vnhsge-2018)

## TÓM TẮT / ABSTRACT

Nghiên cứu này tập trung vào việc phân tích dữ liệu điểm thi Trung học Phổ thông Quốc Gia năm 2018 tại Việt Nam. Dữ liệu được thu thập từ hàng nghìn thí sinh tham gia kỳ thi này và bao gồm thông tin về điểm số từ các môn thi cơ bản như Toán, Văn, Ngoại Ngữ, .... Mục tiêu của nghiên cứu là phân tích sâu hơn về phân phối điểm số, điều chỉnh dữ liệu điểm thi sao cho hợp lý hơn. Phân tích này sử dụng các phương pháp thống kê và trực quan hóa dữ liệu để trình bày các đặc điểm và các mối quan hệ trong dữ liệu điểm thi. Kết quả từ nghiên cứu có thể cung cấp cái nhìn tổng quan về hiện trạng giáo dục hiện nay, cũng như sự tồn tại bất bình đẳng giữa các môn thi ảnh hưởng đến xét tuyển CĐ, ĐH.

**Từ khóa:** Phân tích thống kê    Cơ sở dữ liệu    Phân phối chuẩn    Định lý Giới hạn trung tâm

# NỘI DUNG



## GIỚI THIỆU / INTRODUCTION

Giới thiệu .....	1
------------------	---

## DỮ LIỆU / DATA OVERVIEW

Nguồn dữ liệu .....	2
Mô hình hóa dữ liệu .....	3

## PHƯƠNG PHÁP / METHODOLOGY

Phương pháp .....	4
-------------------	---

## KẾT QUẢ VÀ BÀN LUẬN / RESULTS & DISCUSSION

Kết quả phân tích môn Toán .....	6
Kết quả phân tích môn Tiếng Anh .....	7
Kết quả phân tích môn Ngữ Văn .....	8
Kết quả phân tích môn Hóa Học .....	9
Kết quả phân tích môn Vật Lý .....	10
Kết quả phân tích môn Sinh Học .....	11
Kết quả phân tích môn Địa Lí .....	12
Kết quả phân tích môn Lịch Sử .....	13
Kết quả phân tích môn GD&CD .....	14
Kết quả phân tích khối A00 .....	15
Kết quả phân tích khối A01 .....	16
Kết quả phân tích khối B00 .....	17
Kết quả phân tích khối C00 .....	18
Kết quả phân tích khối D01 .....	19
So sánh điểm thi giữa các môn thi .....	20
So sánh điểm thi giữa các khối thi .....	21

## KẾT LUẬN / CONCLUSIONS

Kết luận .....	22
----------------	----

TÀI LIỆU THAM KHẢO / REFERENCES .....	23
---------------------------------------	----

PHỤ LỤC / APPENDIX .....	24
--------------------------	----

**Kỳ thi THPT Quốc gia ở Việt Nam** là kỳ thi quan trọng nhất dành cho học sinh phổ thông sau khi hoàn thành cấp học trung học phổ thông (THPT). Đây là kỳ thi mang tính quyết định cao đối với việc **xét tốt nghiệp THPT** và là cơ sở chính để **xét tuyển vào các trường đại học, cao đẳng**.

Kết quả của kỳ thi được biểu thị thông qua **phổ điểm**. Đây là hệ thống điểm số của các thí sinh tham gia vào từng môn thi cụ thể. Phổ điểm thường được biểu diễn dưới dạng biểu đồ, với trục hoành ghi điểm số và trục tung ghi số lượng thí sinh đạt điểm tương ứng.

$$\text{ĐXTN} = \frac{\frac{\text{Tổng điểm 4 bài thi THPT Quốc Gia} + \text{Điểm khuyến khích}}{4} + \text{Điểm trung bình cả năm lớp 12}}{2} + \text{Điểm ưu tiên}$$

*Công thức tính điểm xét tốt nghiệp năm 2018, ĐXTN  $\geq$  5.0 được gọi là đạt*

Trong các năm diễn ra kỳ thi THPT Quốc Gia (2015 - 2019) (Trước năm 2015 là 2 kỳ thi riêng và sau 2019 được đổi thành kỳ thi tốt nghiệp), năm **2018** được xem là một năm có nhiều biến động khi là năm thứ 2 các môn thi được chuyển sang hình thức thi trắc nghiệm và được đánh giá là có đề thi khó nhất trong các năm. Trong phạm vi bài viết này, chúng ta sẽ **đánh giá và xem xét phổ điểm thi giữa các môn thi** đồng thời **điều chỉnh phổ điểm khối thi** cho chính xác.

**XÃ HỘI** > Giáo dục

## Kỳ thi THPT Quốc gia 2018: Đề thi khó và dài gây tranh luận

### Thi THPT Quốc gia 2018: Đề Toán quá dài và khó, không phù hợp với kỳ thi THPT

**GIÁO DỤC** | Thứ Sáu, 29/06/2018 11:42:00 +07:00

 Theo dõi **VTC NEWS** trên **Google News**

*Một số bài báo nói về kỳ thi THPT Quốc Gia 2018*

Ngoài ra, phổ điểm khối thi hiện nay trên các trang phương tiện truyền thông chưa phản ánh chính xác vì sau kỳ thi, thí sinh được phép chọn lại khối thi và nguyện vọng xét tuyển, phần lớn thí sinh sẽ chỉ chọn khối thi điểm cao nhất. Việc lọc các khối thi điểm thấp của thí sinh sẽ cho ra phổ điểm chính xác hơn về phân phối điểm.

**Nguồn dữ liệu:** Trích xuất dữ liệu từ API của trang **TTXVN** đã cung cấp với đầu vào là số báo danh của thí sinh.

```
{
  "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"
  }
}
```

**Mẫu truy vấn:** Kết quả trả về dữ liệu ở dạng JSON với điểm thi các môn của thí sinh có SBD "01000111"

```
{
  "result": [
    {
      "CityCode": "01",
      "CityArea": null,
      "Code": "01000111",
      "Toan": "4.20",
      "NguVan": "6.75",
      "NgoaiNgu": "9.20",
      "VatLi": "",
      "HoaHoc": "",
      "SinhHoc": "",
      "KHTN": "",
      "DiaLi": "5.25",
      "LichSu": "4.75",
      "GDGD": "8.00",
      "KHXX": "6.00",
      ...
    }
  ]
}
```

## Tóm tắt quá trình thu thập dữ liệu:

**Ngôn ngữ lập trình:** Python

1. **[Xác định API và endpoints]** Xác định API cần truy cập và endpoints cụ thể để lấy dữ liệu (Xác định cấu trúc và số báo danh lớn nhất của mỗi tỉnh).
2. **[Sử dụng thư viện để gửi yêu cầu API]** Sử dụng thư viện aiohttp để gửi yêu cầu HTTP đến API và nhận phản hồi từ máy chủ.
3. **[Asyncio để xử lý bất đồng bộ]** Sử dụng asyncio để thực hiện các yêu cầu API một cách bất đồng bộ, giúp tăng tốc độ thu thập dữ liệu.
4. **[Xử lý dữ liệu trả về]** Khi nhận được phản hồi từ API, dữ liệu trả về ở dạng JSON. Đồng thời ghi dữ liệu lên file CSV để lưu trữ.
5. **[Gộp dữ liệu từ các đợt thu thập]** Vì quá trình thu thập sẽ mất nhiều thời gian nên sẽ có các đợt thu thập khác nhau. Vì vậy, cần có bước gộp các file CSV bằng thư viện pandas.

## Mô hình hóa dữ liệu: Quản lý điểm thi của các thí sinh tham gia kỳ thi THPT Quốc Gia

Sau khi thu thập dữ liệu, nhận thấy chỉ có thông tin về Mã thí sinh, Mã tỉnh và điểm số các môn thi. Tuy nhiên, để hiển thị đầy đủ thông tin, cần tiếp tục thu thập dữ liệu và xây dựng một cơ sở dữ liệu quan hệ để lưu trữ thông tin hoàn chỉnh và tạo các liên kết giữa môn học và khối thi để phục vụ quá trình phân tích.

### Lược đồ Cơ sở dữ liệu:

**PROVINCE**(PROVINCE\_CODE, PROVINCE\_NAME)

*Tên từ:* Mỗi tỉnh/thành có một mã số và tên tỉnh/thành.

**CANDIDATE**(CANDIDATE\_ID, PROVINCE\_CODE)

*Tên từ:* Mỗi thí sinh tham gia có một mã số dự thi và mã tỉnh để cho biết thí sinh dự thi ở tỉnh/thành nào.

**SUBJECT**(SUBJECT\_ID, SUBJECT\_NAME)

*Tên từ:* Mỗi môn thi chứa một mã môn thi và tên môn thi đó.

**BLOCK**(BLOCK\_ID, BLOCK\_NAME)

*Tên từ:* Mỗi khối thi chứa một mã khối thi và diễn giải khối thi đó.

**SUBJECT\_BLOCK**(BLOCK\_ID, SUBJECT\_ID)

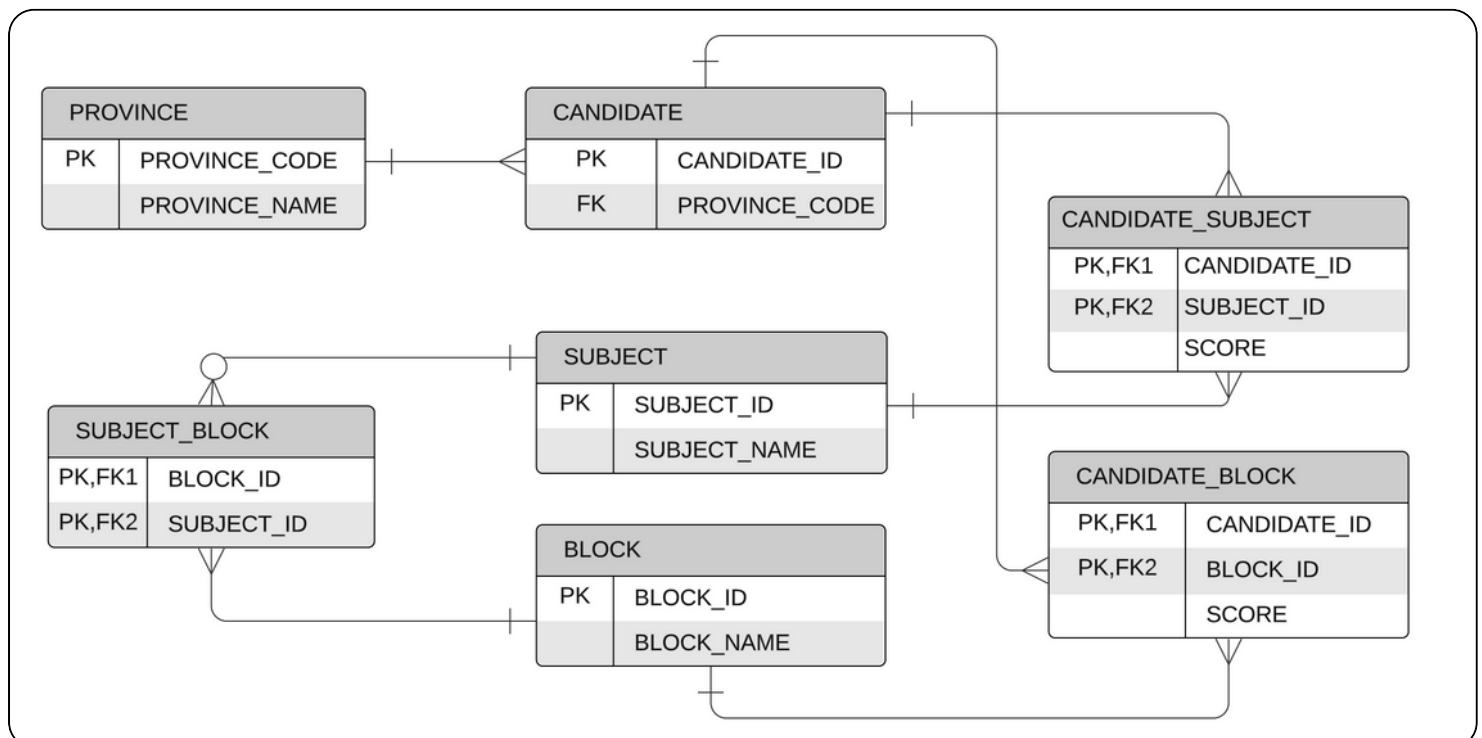
*Tên từ:* Diễn giải chi tiết trong mỗi khối thi bao gồm những môn thi nào.

**CANDIDATE\_SUBJECT**(CANDIDATE\_ID, SUBJECT\_ID, SCORE)

*Tên từ:* Diễn giải chi tiết mỗi thí sinh tham gia những môn thi nào và điểm thi của mỗi môn thi.

**CANDIDATE\_BLOCK**(CANDIDATE\_ID, BLOCK\_ID, SCORE)

*Tên từ:* Diễn giải chi tiết mỗi thí sinh có các khối thi nào được hình thành và điểm thi của mỗi khối thi.



Entity-Relationship Diagram (Crow's foot notation)

Lọc ảo điểm của môn thi và khối thi (Outliers)

Vì thông tin về số lượng và chi tiết về thí sinh tham gia không được cung cấp đầy đủ, không thể xác định liệu thí sinh tham gia để **xét tốt nghiệp** hay để **xét tuyển vào các trường Cao đẳng, Đại học**. Do đó, chúng ta có thể **giả định** rằng **thí sinh có tổng điểm cao nhất trong một khối thi nào thì ba môn thi trong khối đó sẽ được coi là các môn thi chính và khối thi chính**.

Mặc dù theo quy định, thí sinh tham gia ít nhất 6 môn thi, nhưng hầu hết thí sinh thường học lệch và chỉ tập trung vào 1 hoặc 2 khối thi (từ 3 đến 4 môn thi) để xét tuyển vào đại học, các môn thi khác chỉ học ở mức chống liệt và với hình thức thi tự trắc nghiệm, điểm của các nhóm này thường rơi vào 2-4 điểm, điều này làm phản ánh không đúng chính xác phổ điểm. Theo báo cáo, có **90% thí sinh chọn các khối thi truyền thống** (A00, A01, B00, C00, D01) nên chúng ta sẽ sử dụng các khối này để xét.

Tương quan môn thi và khối thi		A00	A01	B00	C00	D01
Nhóm môn bắt buộc	Toán	✓	✓	✓		✓
	Ngữ Văn				✓	✓
	Ngoại Ngữ		✓			✓
Nhóm môn Khoa học Tự nhiên	Vật Lý	✓	✓			
	Hóa Học	✓		✓		
	Sinh Học			✓		
Nhóm môn Khoa học Xã hội	Lịch Sử				✓	
	Địa Lý				✓	
	GD&CD					

Định lý Giới hạn trung tâm (Central Limit Theorem)

Ngoài lý do chủ quan trên thì còn có một lý do khách quan là ảnh hưởng của **CLT** lên phổ điểm khối thi hiện nay. Theo CLT, khi tổng của một số lượng lớn các biến ngẫu nhiên độc lập có cùng phân phối (không cần thiết phải chuẩn) thì phân phối của tổng này sẽ gần với phân phối chuẩn, ngay cả khi các biến này có phân phối khác nhau.

Xét điểm thi của 3 môn thi trong một khối thi là các biến ngẫu nhiên độc lập, tổng điểm của 3 môn thi luôn hình thành theo phân phối chuẩn hoặc gần chuẩn do số lượng thí sinh lớn, vì vậy việc **giảm số lượng thí sinh có hình thành khối thi không phải là khối để xét tuyển đại học** sẽ cho ra một phổ điểm chính xác hơn về mức độ ảnh hưởng của điểm thi các môn lên phổ điểm.

Thống kê mô tả (Descriptive Statistical Measures)

Sử dụng một số giá trị đo lường mức độ tập trung của dữ liệu như **Mean** (trung bình), **Mode** (yếu vị) và **Median** (Trung vị). Khi so sánh ba giá trị này, chúng ta có thể có cái nhìn tổng quan về trung bình điểm, điểm xuất hiện nhiều nhất và điểm ở vị trí trung tâm của phân phối điểm thi. Điều này có thể giúp chúng ta hiểu rõ hơn về đặc điểm chung của bài thi.

## Đánh giá phổ điểm bằng phân phối chuẩn (Gaussian Distribution)

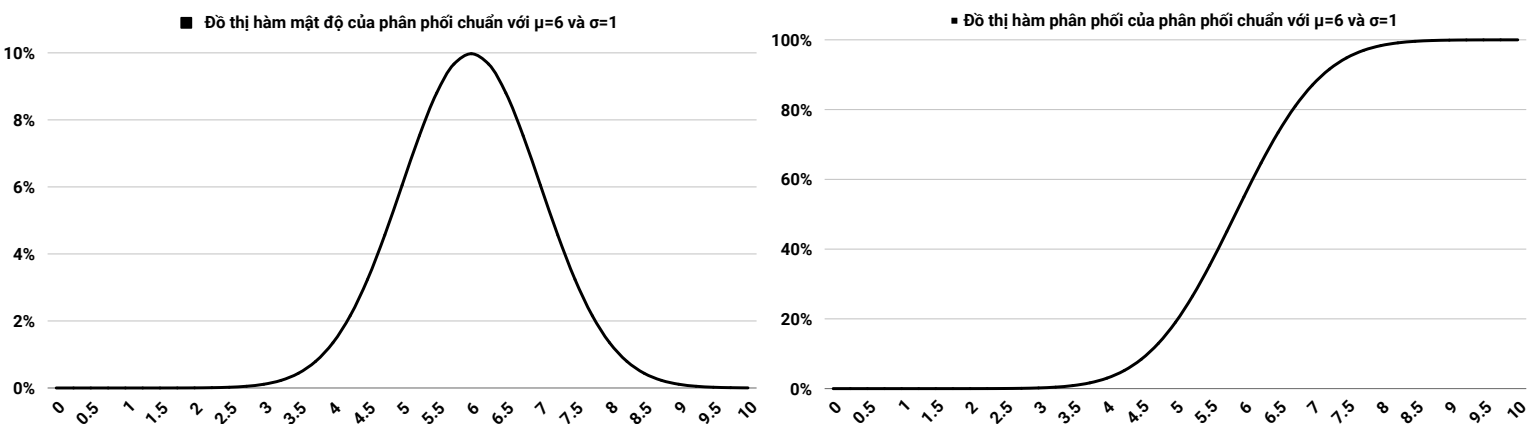
Nếu xét về cách tính điểm và mục tiêu của kỳ thi trung học phổ thông (THPT) và kỳ thi đại học, chúng có **mục đích khác nhau** trong việc đánh giá và phân loại học sinh theo các chuẩn đầu ra cụ thể.

- Để xét tốt nghiệp, phổ điểm sẽ có hình chữ J (**CRT**).
- Để xét tuyển đại học, phổ điểm có hình quả chuông (**NRT**) để phân loại trình độ học sinh.

Theo công thức tính điểm thi xét tốt nghiệp, trung bình điểm thi của các môn thi phải đạt tối thiểu là 5.0 (trường hợp điểm trung bình môn học là 5.0) nên đây sẽ là **ngưỡng tiêu chuẩn** và phần lớn thí sinh phải vượt qua.

Đối với điểm thi dùng cho xét tuyển Đại học, mức điểm sẽ dùng trong khoảng 6 - 10 điểm và đảm bảo có sự phân hóa rõ ràng giữa các nhóm điểm.

Kết hợp hai phần này, phổ điểm tổng thể sẽ có dạng hình quả chuông, với "đỉnh chuông" tập trung từ 5.5 đến 6.0 điểm. Trong đó, khoảng điểm từ 5.0 đến 6.0 được áp dụng cho phần thi THPT và từ 6.0 đến 10.0 điểm áp dụng cho việc phân loại thí sinh thi đại học.



Phổ điểm lý tưởng cho kỳ thi THPT Quốc Gia 2018

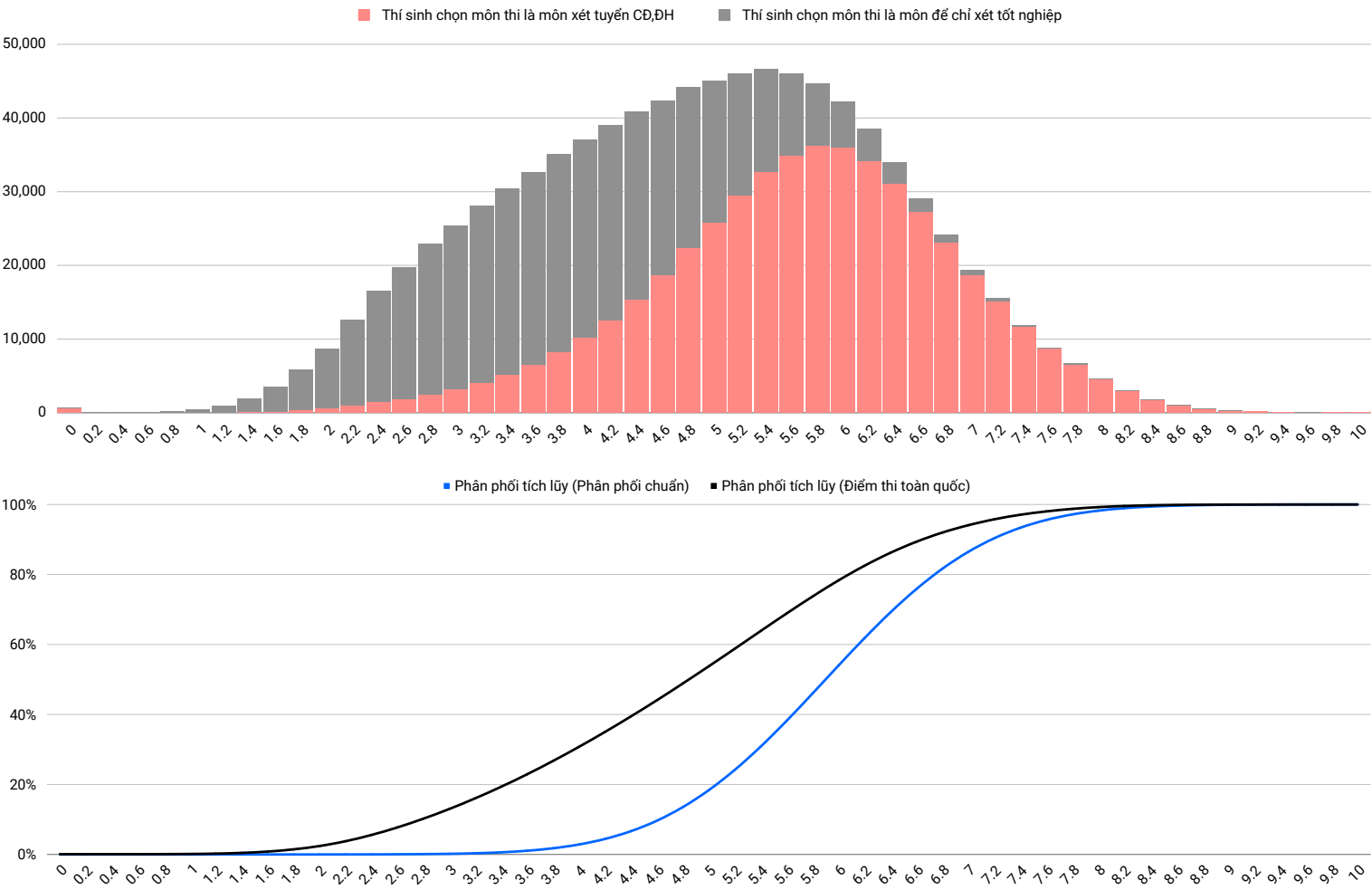
Từ kết luận trên, phân phối chuẩn được chọn có điểm trung bình ( $\mu$ ) là 6.0 và độ lệch chuẩn ( $\sigma$ ) là 1.0. Theo quy tắc 3 $\sigma$  (**Three-Sigma Limits**), 99.73% điểm thí sinh sẽ nằm trong khoảng ( $\mu - 3\sigma, \mu + 3\sigma$ ) ~ (3,9).

Hạn chế "mưa điểm 10" là một trong những tiêu chí quan trọng trong kỳ thi. Trong năm 2017, có đến 4,153 điểm 10 và làm một số ngành có tỉ lệ chọi cao có mức điểm vượt ngưỡng 30 và phải cần tới các tiêu chí phụ. Việc đề thi có độ phân hóa thấp có thể làm ảnh hưởng đến chất lượng đầu vào trong tuyển sinh.

**CRT (criterion-referenced test): Bài thi đánh giá theo chuẩn**, được dùng để đánh giá người được kiểm tra xem có đáp ứng được những kiến thức, kỹ năng hay năng lực cụ thể theo tiêu chuẩn đã đề ra hay không.

**NRT (Norm-referenced test): Bài thi đánh giá tương đối**, được dùng để so sánh, đánh giá, xếp hạng các thí sinh trong một nhóm chuẩn mực cụ thể.

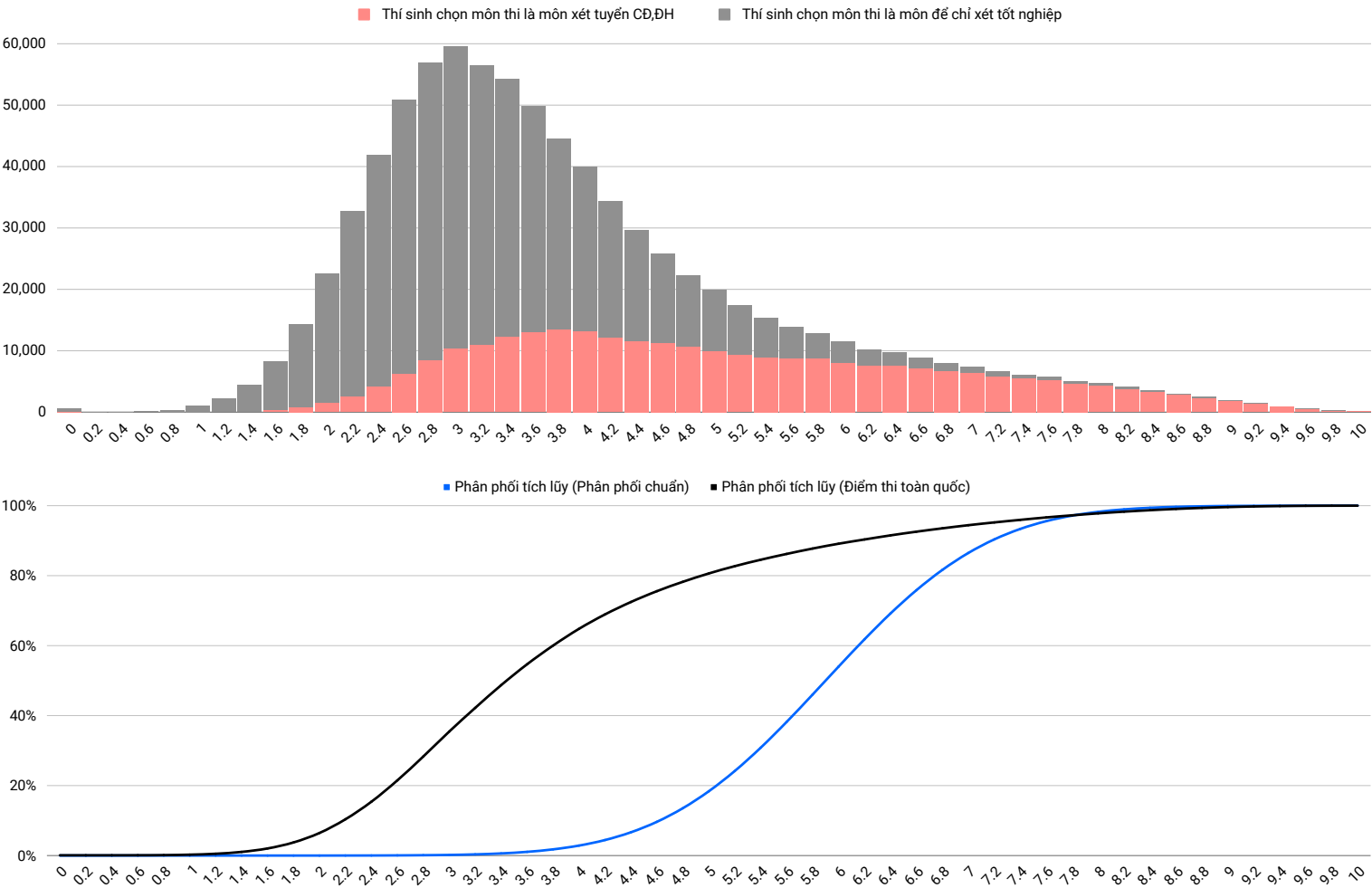
Môn thi: Toán (50 câu trắc nghiệm)



Mục	Phân phối chuẩn $\mu=6$ và $\sigma=1$	Thí sinh toàn quốc	Thí sinh chọn môn Toán là môn xét tuyển
Tổng số bài thi	919,139	919,139	499,600 (chiếm 54%)
Mức điểm cao nhất	10	10	10
Số lượng bài có mức điểm cao nhất	25	2	2
Số lượng bài thi đạt mức điểm giỏi ( $\geq 8$ )	26,186 (chiếm 2.85%)	11,560 (chiếm 1.258%)	11,520 (chiếm 2.306%)
Số lượng bài thi < 5	124,331 (chiếm 13.53%)	449,127 (chiếm 48.864%)	115,875 (chiếm 23.194%)
Điểm trung bình   Mean	6.0	4.88	5.69
Điểm trung vị   Median	6.0	4.8	5.6
Điểm yếu vị   Mode	6.0	5.4	5.8

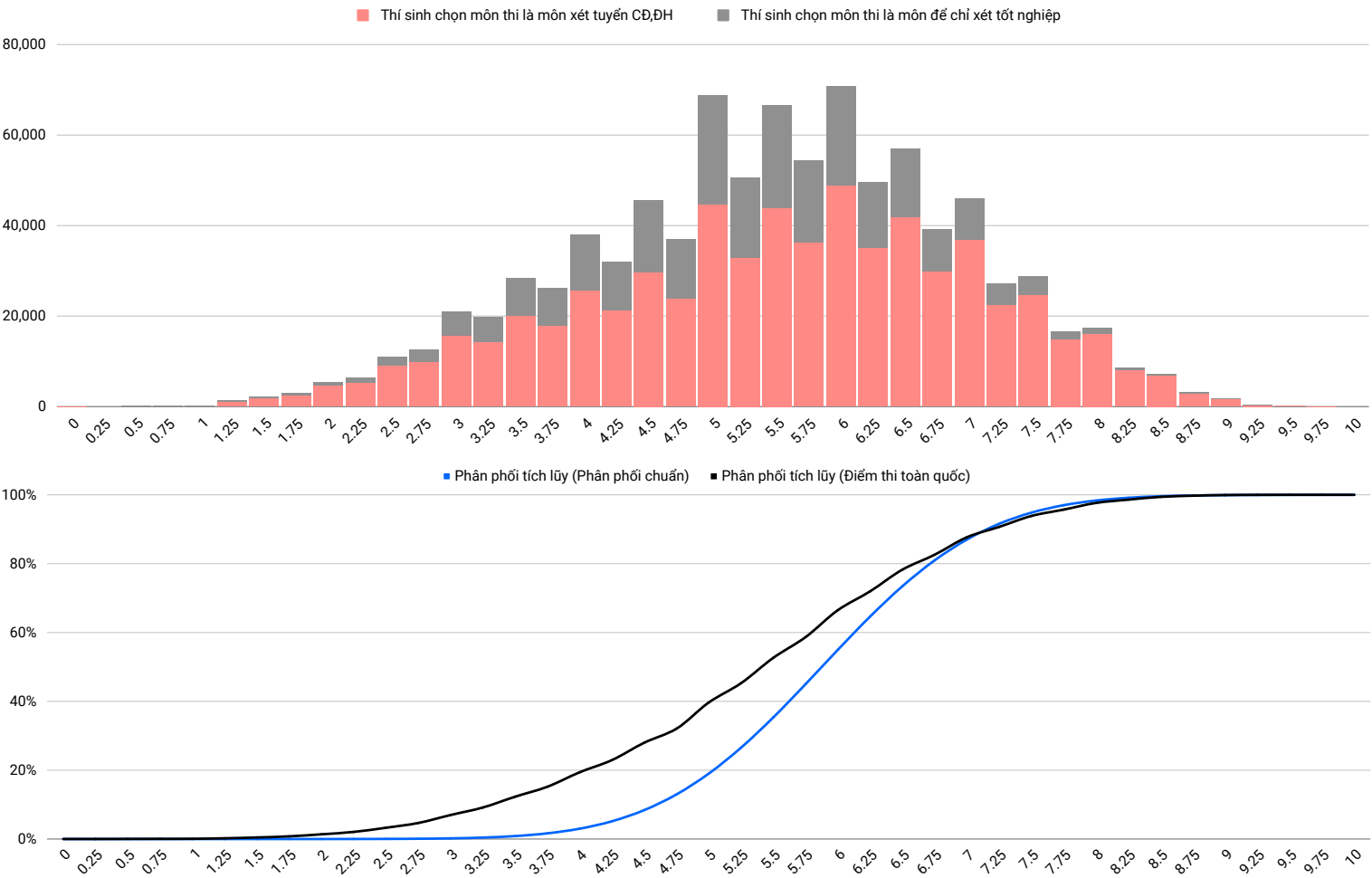


Môn thi: Tiếng Anh (50 câu trắc nghiệm)



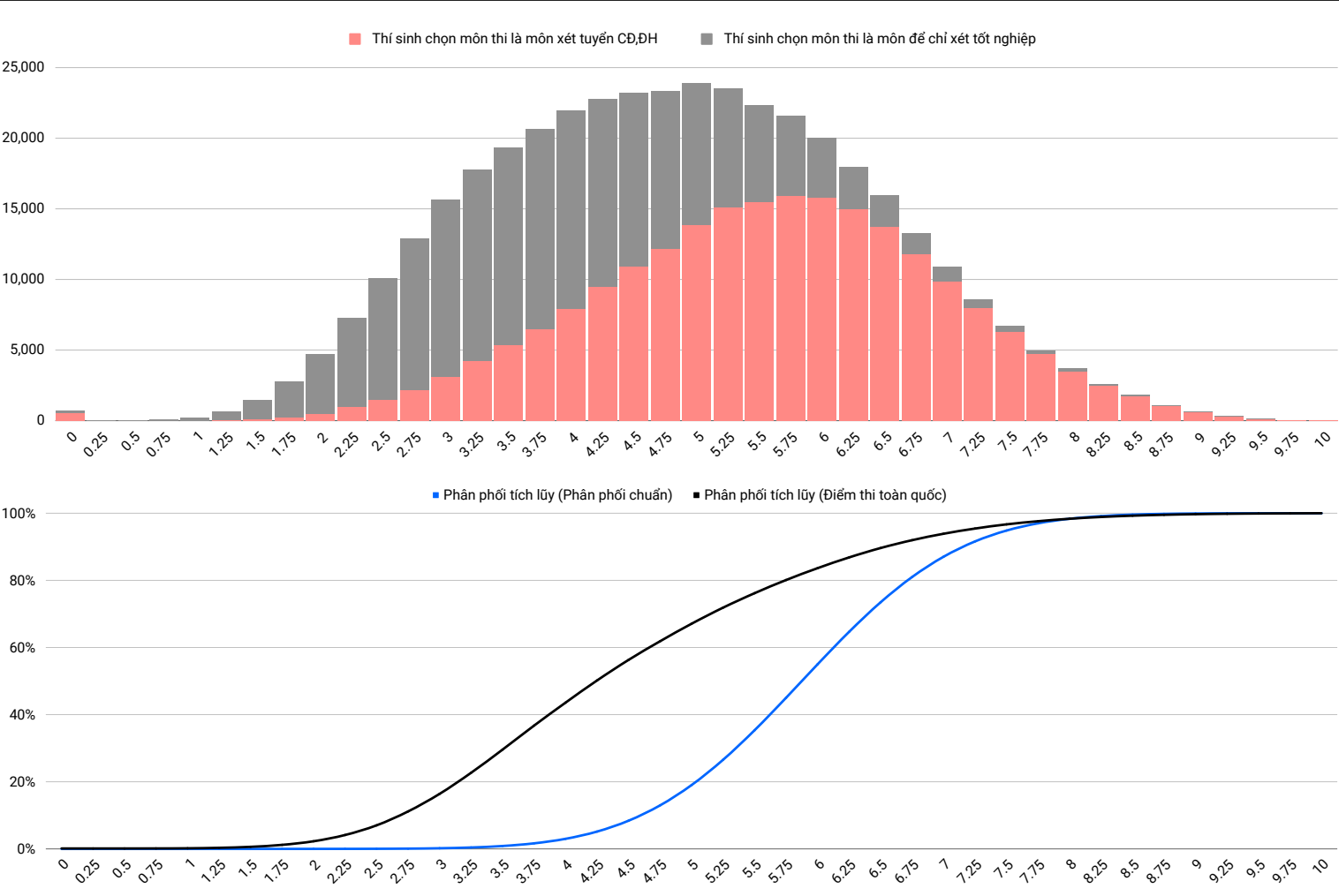
Mục	Kỳ vọng	Thí sinh toàn quốc	Thí sinh chọn môn Tiếng Anh là môn xét tuyển
Tổng số bài thi	834,690	834,690	277,689 (chiếm 33%)
Mức điểm cao nhất	10	10	10
Số lượng bài có mức điểm cao nhất	22	106	106
Số lượng bài thi đạt mức điểm giỏi (>=8)	23,780 (chiếm 2.85%)	23,035 (chiếm 2.76%)	22,129 (chiếm 7.969%)
Số lượng bài thi < 5	112,907 (chiếm 13,53%)	653,016 (chiếm 78.235%)	144,530 (chiếm 52.047%)
Điểm trung bình   Mean	6.0	3.92	5.07
Điểm trung vị   Median	6.0	3.4	4.6
Điểm yếu vị   Mode	6.0	3.0	3.8

Môn thi: Ngữ Văn (Hình thức thi tự luận)



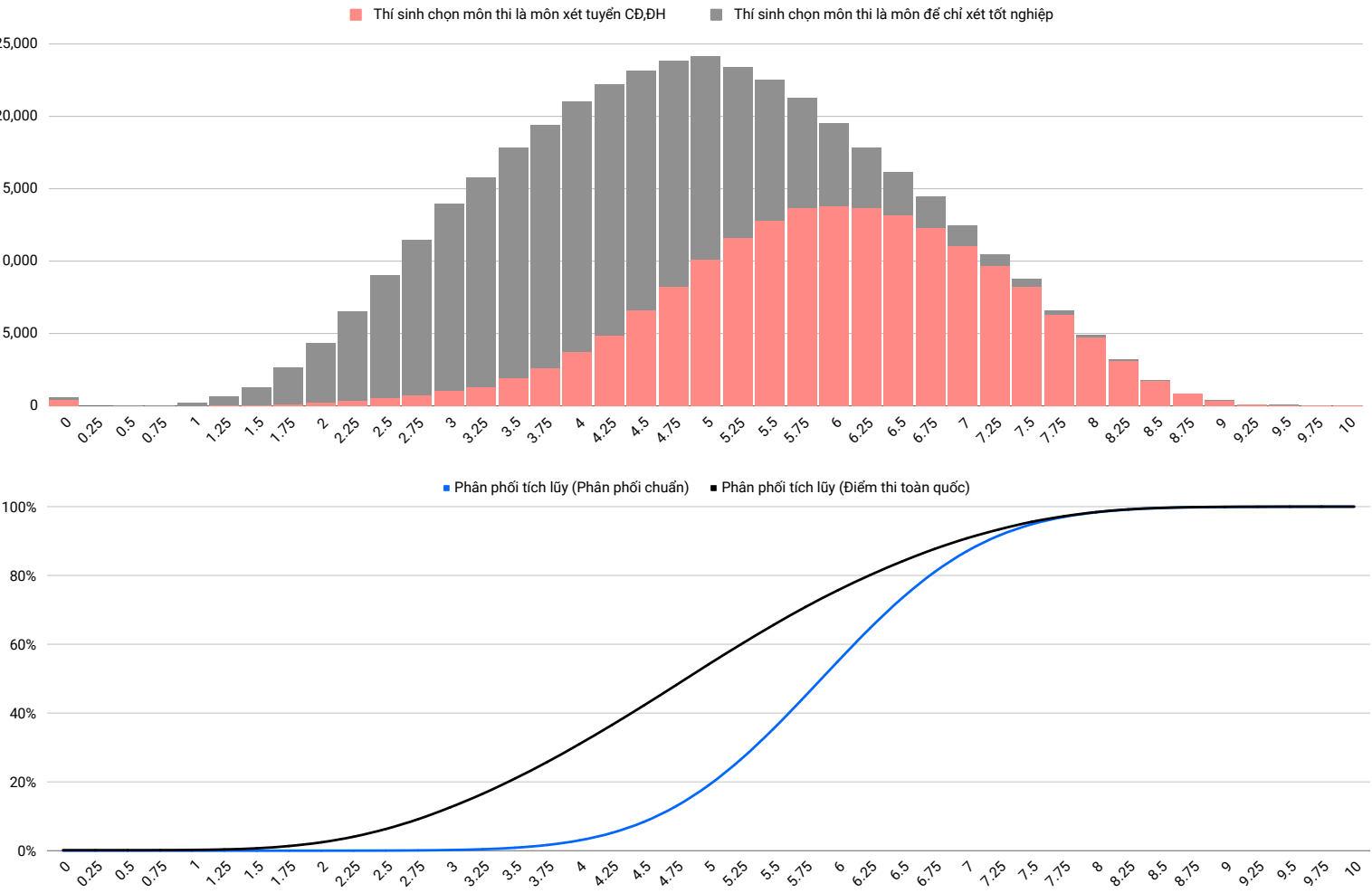
Mục	Kỳ vọng	Thí sinh toàn quốc	Thí sinh chọn môn Ngữ Văn là môn xét tuyển
Tổng số bài thi	903,659	903,659	654, 216 (chiếm 72%)
Mức điểm cao nhất	10	9.75	9.75
Số lượng bài có mức điểm cao nhất	30	7	7
Số lượng bài thi đạt mức điểm giỏi (>=8)	27,149 (chiếm 3.004%)	38,684 (chiếm 4.281%)	36,412 (chiếm 5.566%)
Số lượng bài thi < 5	117,181 (chiếm 12.967%)	289,745 (chiếm 32.064%)	204,551 (chiếm 31.267%)
Điểm trung bình   Mean	6.0	5.45	5.53
Điểm trung vị   Median	6.0	5.25	5.5
Điểm yếu vị   Mode	6.0	6.0	6.0

Môn thi: Hóa Học (40 câu trắc nghiệm)



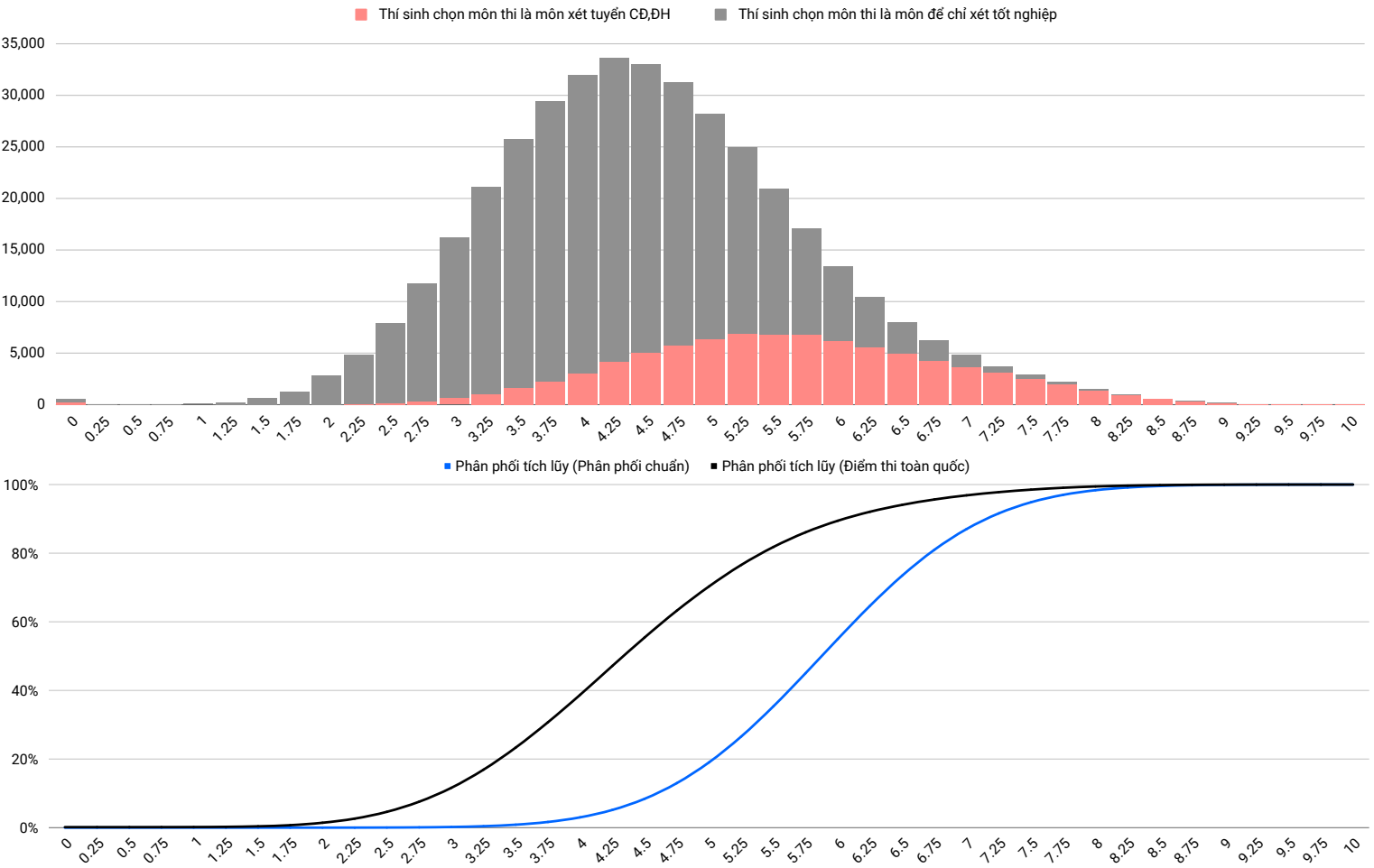
Mục	Kỳ vọng	Thí sinh toàn quốc	Thí sinh chọn môn Hóa Học là môn xét tuyển
Tổng số bài thi	405,859	405,859	221,910 (chiếm 55%)
Mức điểm cao nhất	10	10	10
Số lượng bài có mức điểm cao nhất	14	15	15
Số lượng bài thi đạt mức điểm giỏi (>=8)	12,193 (chiếm 3.004%)	10,333 (chiếm 2.546%)	9,972 (chiếm 4.494%)
Số lượng bài thi < 5	117,181 (chiếm 12.967%)	289,745 (chiếm 32.064%)	204,551 (chiếm 31.267%)
Điểm trung bình   Mean	6.0	4.86	5.58
Điểm trung vị   Median	6.0	4.5	5.5
Điểm yếu vị   Mode	6.0	5.9	5.75

Môn thi: Vật Lý (40 câu trắc nghiệm)



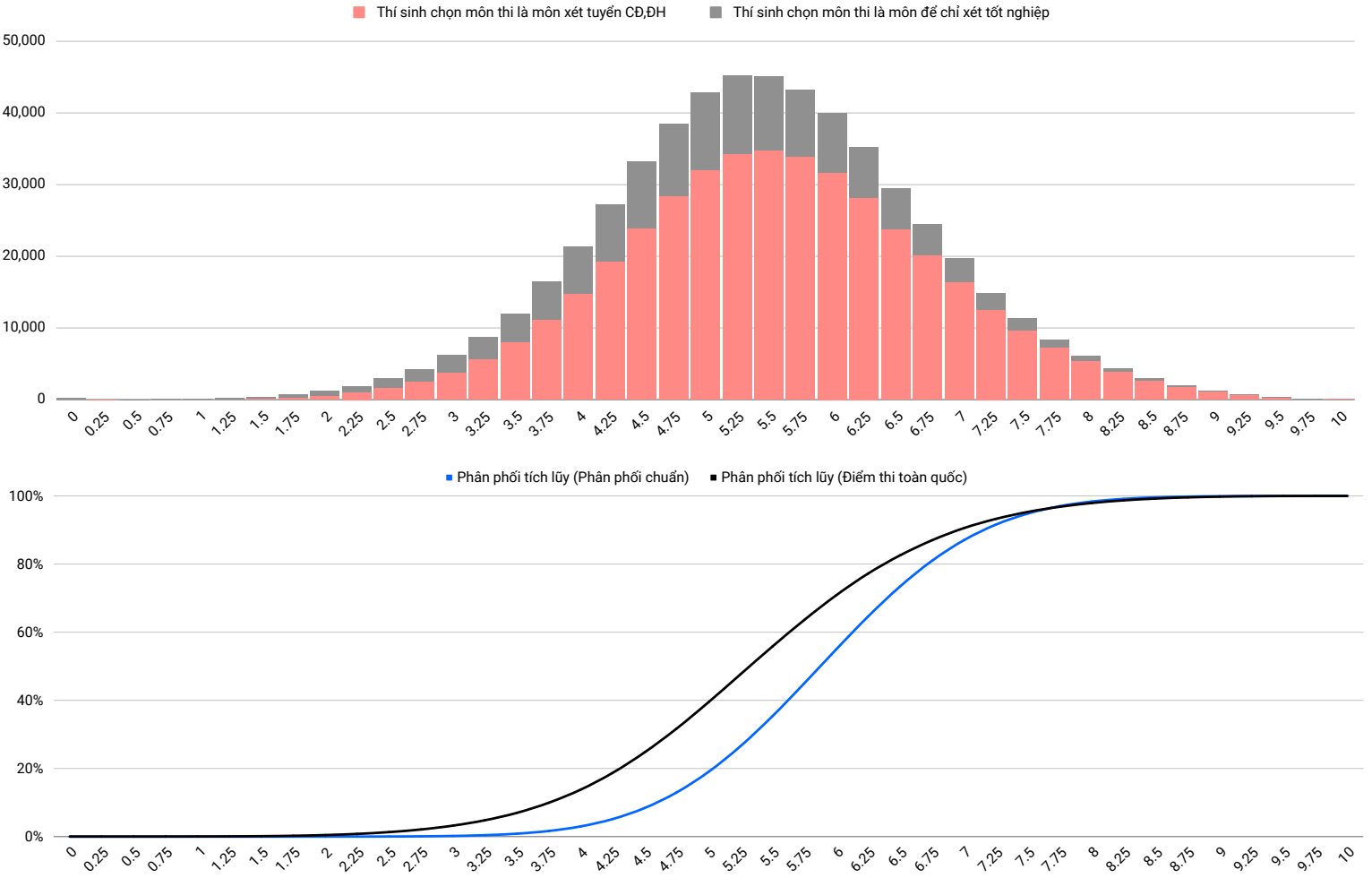
Mục	Kỳ vọng	Thí sinh toàn quốc	Thí sinh chọn môn Vật Lý là môn xét tuyển
Tổng số bài thi	403,167	403,167	180,947 (chiếm 45%)
Mức điểm cao nhất	10	10	10
Số lượng bài có mức điểm cao nhất	13	3	3
Số lượng bài thi đạt mức điểm giỏi (>=8)	12,113 (chiếm 3.004%)	11,307 (chiếm 2.805%)	11,127 (chiếm 6.149%)
Số lượng bài thi < 5	52,280 (chiếm 12.967%)	289,745 (chiếm 48.175%)	204,551 (chiếm 18.311%)
Điểm trung bình   Mean	6.0	4.97	5.99
Điểm trung vị   Median	6.0	4.75	5.75
Điểm yếu vị   Mode	6.0	5.0	6.0

Môn thi: Sinh Học (40 câu trắc nghiệm)



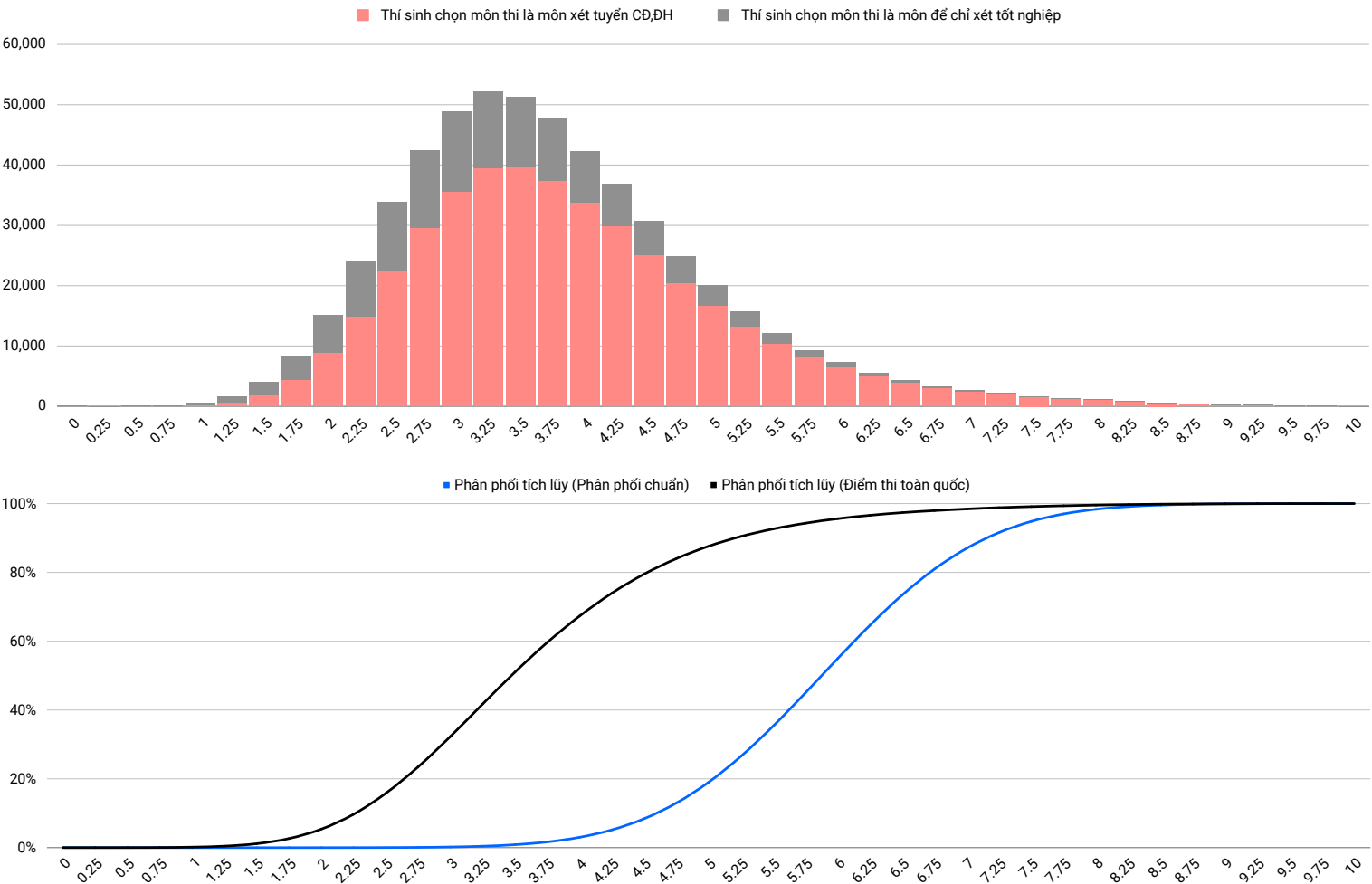
Mục	Kỳ vọng	Thí sinh toàn quốc	Thí sinh chọn môn Sinh Học là môn xét tuyển
Tổng số bài thi	399,728	399,728	87,694 (chiếm 22%)
Mức điểm cao nhất	10	10	10
Số lượng bài có mức điểm cao nhất	10	2	2
Số lượng bài thi đạt mức điểm giỏi (>=8)	9,375 (chiếm 3.004%)	3,749 (chiếm 0.938%)	3609 (chiếm 4.115%)
Số lượng bài thi < 5	40,463 (chiếm 12.967%)	228,021 (chiếm 63.228%)	24,721 (chiếm 28.190%)
Điểm trung bình   Mean	6.0	4.54	5.63
Điểm trung vị   Median	6.0	4.25	5.25
Điểm yếu vị   Mode	6.0	4.25	5.25

Môn thi: Địa Lý (40 câu trắc nghiệm)



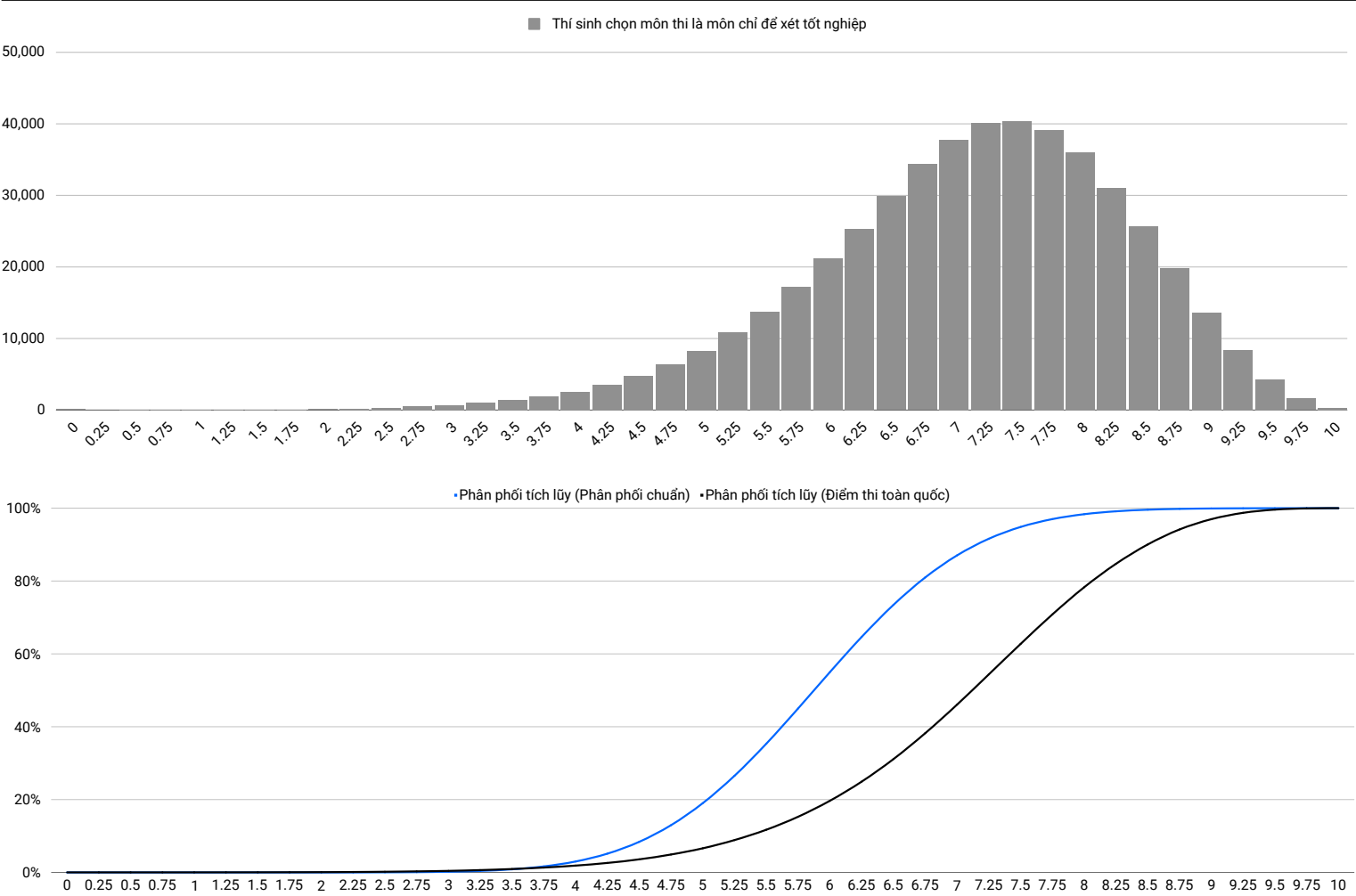
Mục	Kỳ vọng	Thí sinh toàn quốc	Thí sinh chọn môn Địa Lý là môn xét tuyển
Tổng số bài thi	552,492	552,492	423,257 (chiếm 77%)
Mức điểm cao nhất	10	10	10
Số lượng bài có mức điểm cao nhất	18	32	32
Số lượng bài thi đạt mức điểm giỏi (>=8)	16,599 (chiếm 3.004%)	17,557 (chiếm 3.178%)	16,259 (chiếm 3.841%)
Số lượng bài thi < 5	71,644 (chiếm 12.967%)	175,203 (chiếm 31.711%)	121,937 (chiếm 28.809%)
Điểm trung bình   Mean	6.0	5.46	5.57
Điểm trung vị   Median	6.0	5.25	5.25
Điểm yếu vị   Mode	6.0	5.25	5.5

Môn thi: Lịch Sử (40 câu trắc nghiệm)



Mục	Kỳ vọng	Thí sinh toàn quốc	Thí sinh chọn môn Lịch Sử là môn xét tuyển
Tổng số bài thi	553,056	552,492	423,257 (chiếm 77%)
Mức điểm cao nhất	10	10	10
Số lượng bài có mức điểm cao nhất	19	10	10
Số lượng bài thi đạt mức điểm giỏi (>=8)	16,616 (chiếm 3.004%)	3,486 (chiếm 0.63%)	3,405 (chiếm 0.804%)
Số lượng bài thi < 5	71717 (chiếm 12.967%)	464,595 (chiếm 84.005%)	345,284 (chiếm 81.578%)
Điểm trung bình   Mean	6.0	3.76	3.89
Điểm trung vị   Median	6.0	3.5	3.25
Điểm yếu vị   Mode	6.0	3.5	3.25

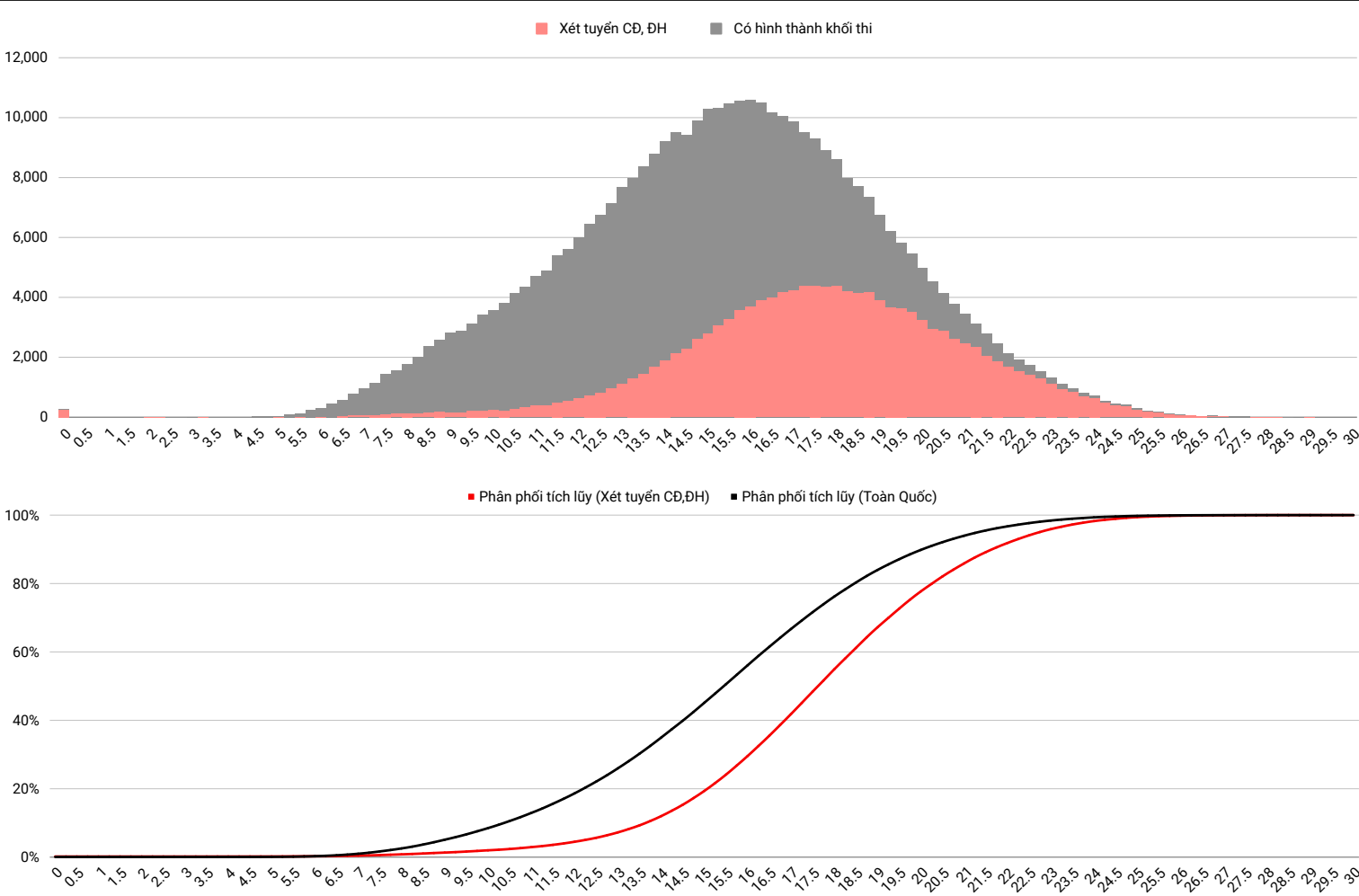
Môn thi: GDCD (40 câu trắc nghiệm)



Mục	Kỳ vọng	Thí sinh toàn quốc
Tổng số bài thi	482,763	482,763
Mức điểm cao nhất	10	10
Số lượng bài có mức điểm cao nhất	16	309
Số lượng bài thi đạt mức điểm giỏi (>=8)	14,504 (chiếm 3.004%)	140,808 (chiếm 29.167%)
Số lượng bài thi < 5	91,806 (chiếm 12.967%)	23,772 (chiếm 4.924%)
Điểm trung bình   Mean	6.0	7.12
Điểm trung vị   Median	6.0	7.0
Điểm yếu vị   Mode	6.0	7.5

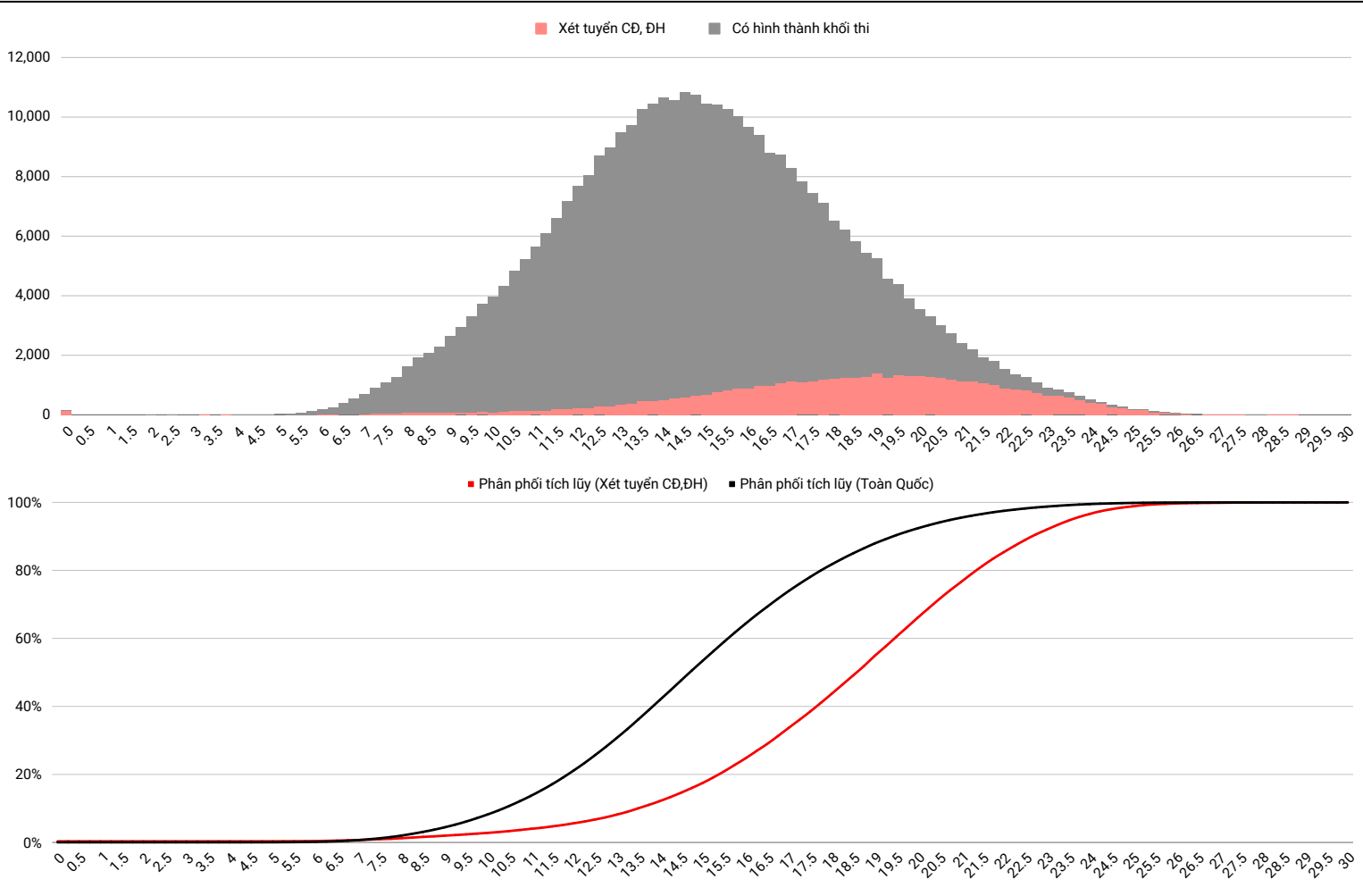


Khối thi: A00 (Toán - Vật Lý - Hóa Học)



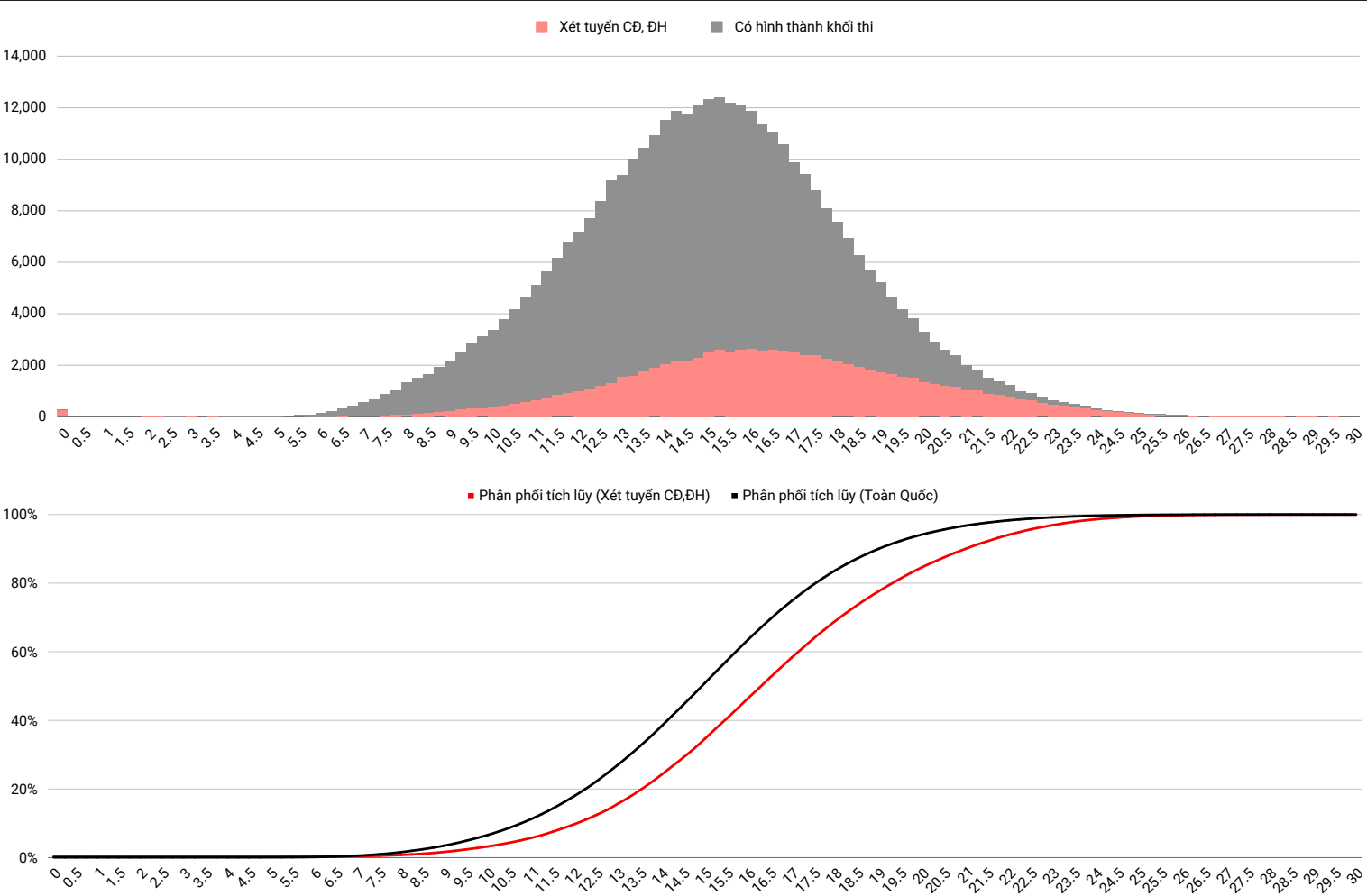
Mục	Thí sinh Xét tuyển CĐ, ĐH	Thí sinh toàn quốc
Tổng số bài thi	134,217 (chiếm 35%)	383,008
Mức điểm cao nhất	29	29
Số lượng bài có mức điểm cao nhất	1	1
Số lượng bài thi đạt điểm $\geq 24.0$	2,993 (chiếm 2.230%)	3,258 (chiếm 0.851%)
Số lượng bài thi < 15	23,198 (chiếm 17.284%)	162,684 (chiếm 42.475%)
Điểm trung bình   Mean	17.68	15.46
Điểm trung vị   Median	17.5	15.25
Điểm yếu vị   Mode	18.0	16.0

Khối thi: A01 (Toán - Vật Lý - Tiếng Anh)



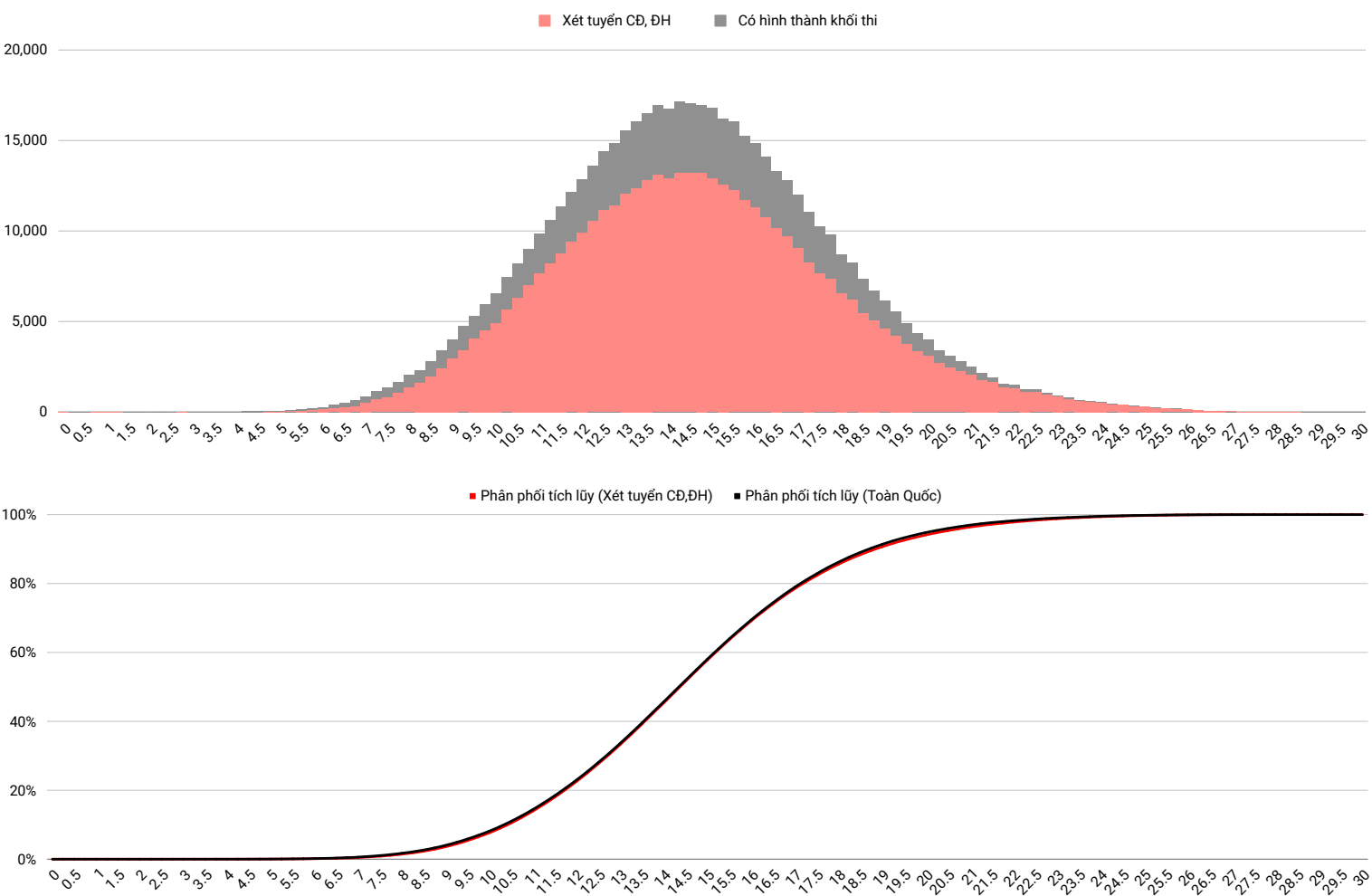
Mục	Thí sinh Xét tuyển CĐ, ĐH	Thí sinh toàn quốc
Tổng số bài thi	46,730 (chiếm 13%)	368,666
Mức điểm cao nhất	28.75	28.75
Số lượng bài có mức điểm cao nhất	2	2
Số lượng bài thi đạt điểm $\geq 24.0$	2,013 (chiếm 4.308%)	2,359 (chiếm 0.64%)
Số lượng bài thi $< 15$	7,490 (chiếm 16.028%)	186,442 (chiếm 50.572%)
Điểm trung bình   Mean	18.34	15.46
Điểm trung vị   Median	18.5	15.25
Điểm yếu vị   Mode	19.0	16.0

Khối thi: B00 (Toán - Hóa Học - Sinh Học)



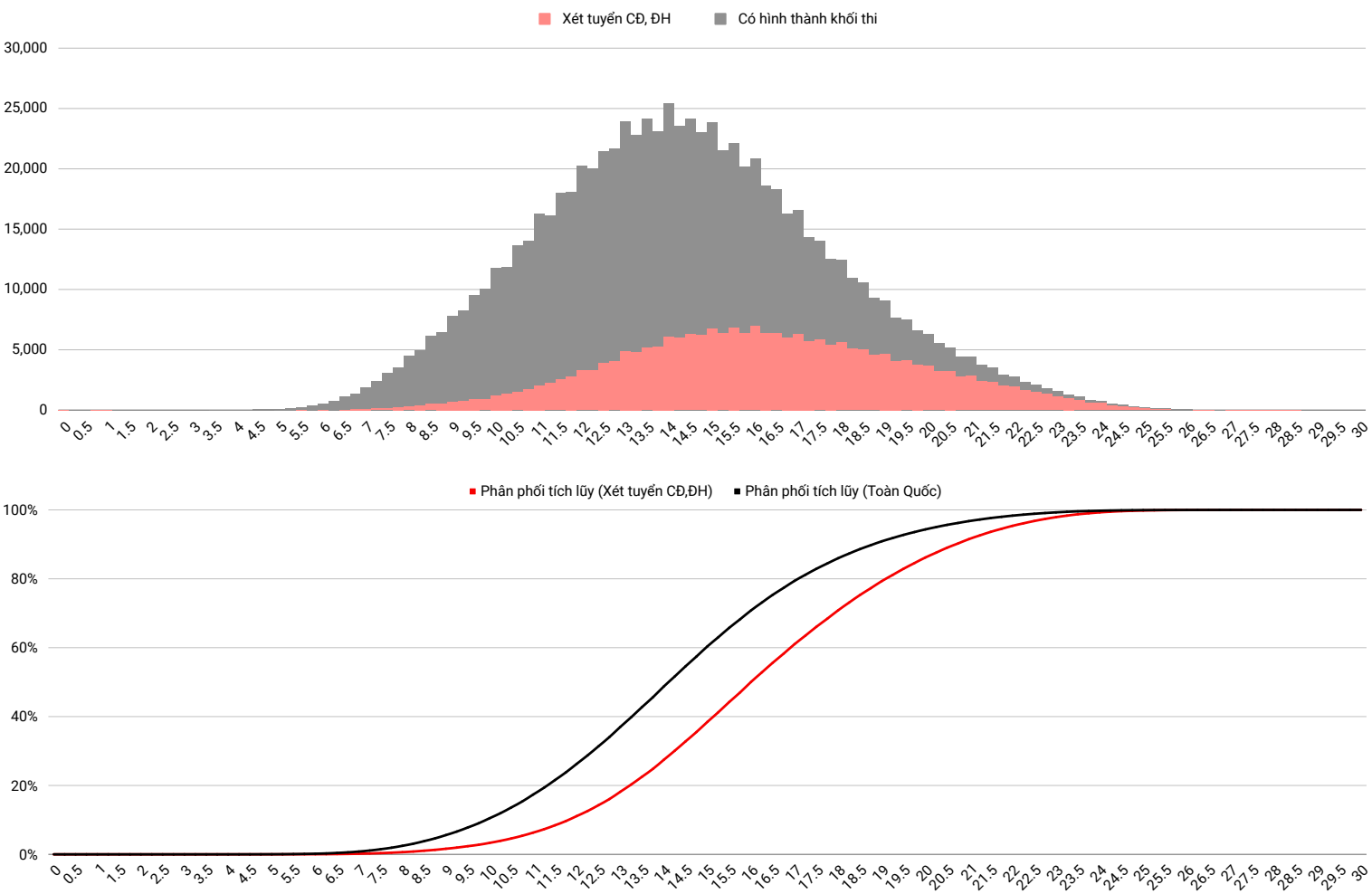
Mục	Thí sinh Xét tuyển CĐ, ĐH	Thí sinh toàn quốc
Tổng số bài thi	87,694 (chiếm 23%)	382,144
Mức điểm cao nhất	29.5	29.5
Số lượng bài có mức điểm cao nhất	1	1
Số lượng bài thi đạt điểm $\geq 24.0$	1,509 (chiếm 1.721%)	1,629 (chiếm 0.426%)
Số lượng bài thi < 15	27,838 (chiếm 31.744%)	182,083 (chiếm 47.648%)
Điểm trung bình   Mean	16.50	15.04
Điểm trung vị   Median	16.25	14.75
Điểm yếu vị   Mode	16	15.25

Khối thi: C00 (Ngữ Văn - Lịch Sử - Địa Lý)



Mục	Thí sinh Xét tuyển CD, ĐH	Thí sinh toàn quốc
Tổng số bài thi	423,257 (chiếm 77%)	548,558
Mức điểm cao nhất	28.5	28.5
Số lượng bài có mức điểm cao nhất	2	2
Số lượng bài thi đạt điểm $\geq 24.0$	3,116 (chiếm 0.736%)	3,146 (chiếm 0.574%)
Số lượng bài thi $< 15$	231,664 (chiếm 54.734%)	301,791 (chiếm 55.015%)
Điểm trung bình   Mean	14.66	14.59
Điểm trung vị   Median	14.25	14.25
Điểm yếu vị   Mode	14.25	14.25

Khối thi: D01 (Toán - Ngữ Văn - Tiếng Anh)



Mục	Thí sinh Xét tuyển CĐ, ĐH	Thí sinh toàn quốc
Tổng số bài thi	230,959 (chiếm 28%)	811,965
Mức điểm cao nhất	28.5	28.5
Số lượng bài có mức điểm cao nhất	2	2
Số lượng bài thi đạt điểm $\geq 24.0$	2,276 (chiếm 0.985%)	2,605 (chiếm 0.321%)
Số lượng bài thi $< 15$	82,063 (chiếm 35.531%)	466,503 (chiếm 57.454%)
Điểm trung bình   Mean	16.24	14.38
Điểm trung vị   Median	15.75	14
Điểm yếu vị   Mode	16	14

## Mức độ khó - dễ của đề thi được phản ánh thông qua phổ điểm

		Nhóm môn bắt buộc			Nhóm môn KHTN			Nhóm môn KHXH		
%	PPC	Toán	Anh	Văn	Hóa	Lý	Sinh	Sử	Địa	GDCD
1%	8.25	7.8	8.6	8.5	8.25	8.25	7.75	7.25	8.5	8.25
5%	7.5	7.4	7.2	7.75	7.25	7.5	6.75	6.0	7.5	7.5
10%	7.25	7.0	6.2	7.25	6.75	7.0	6.0	5.25	7.0	7.0
15%	7.0	6.8	5.4	7.0	6.5	6.5	5.75	4.75	6.75	6.5
20%	6.75	6.6	5.0	6.5	6.25	6.25	5.5	4.5	6.5	6.25
30%	6.5	6.2	4.4	6.25	5.75	5.75	5.0	4.0	6.0	5.75

Bảng phân bố điểm thi của các môn thi (nhóm 30% thí sinh có điểm thi cao nhất)

### Phần lớn thí sinh có điểm thi dưới trung bình

Trừ môn **GDCD** đặc thù là môn thi tốt nghiệp nên điểm thi cao chỉ có khoảng **5%** thí sinh dưới điểm trung bình. Các môn **Văn, Hóa, Địa** có tỷ lệ thí sinh dưới trung bình nằm trong khoảng **30-32%**, các môn **Toán, Lý** có tỷ lệ thí sinh dưới trung bình trong khoảng **48%**, còn các môn **Sinh, Tiếng Anh** và **Lịch sử** có tỷ lệ thí sinh dưới trung bình nhiều nhất trong khoảng **63%-84%**.

Nếu chỉ xét các thí sinh chọn môn thi là môn xét tuyển thì tỷ lệ có giảm như môn **Toán, Lý** giảm còn **18%-23%**, môn **Sinh** là **28%**, môn **Tiếng Anh** là **52%**, các môn còn lại có giảm ít hoặc giữ nguyên (**Văn - 31%, Hóa - 31%, Địa - 28%, Sử - 81%**)

### Tiếng Anh và Lịch Sử có điểm trung bình thấp nhưng tính chất khác nhau

Về môn **Tiếng Anh**, do phổ điểm thi khác nhau ở từng địa phương, điểm trung bình cao nhất là **TP.HCM** với **5.1**, thấp nhất là **Hà Giang** với **3.0** (chênh lệch 2.1), các thí sinh đạt mức điểm cao ( $\geq 8$ ) phần lớn là ở các thành phố lớn như Hà Nội (5.800 bài thi), **TP. HCM** (5.400 bài thi), **Hải Phòng** (780 bài thi). Thấp nhất là **Lai Châu** (12 bài thi).

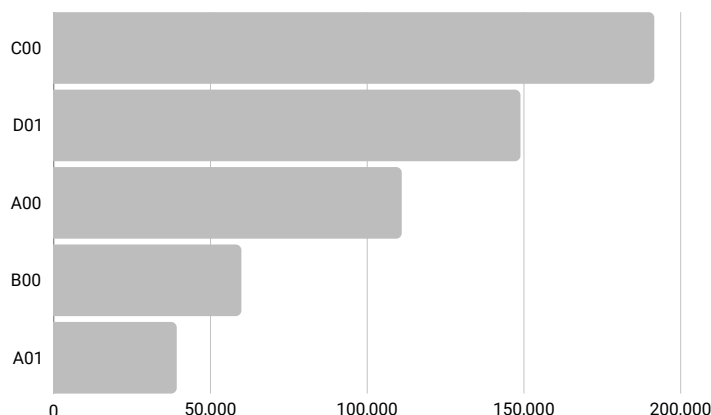
Còn với môn **Lịch Sử**, điểm trung bình cao nhất là **Bạc Liêu** với **4.2**, thấp nhất là **Hà Giang** với **3.5** (chênh lệch 0.7). **Hà Nội** có số bài thi đạt điểm giỏi nhiều nhất với 380 bài thi, thấp nhất là **Trà Vinh** với 11 bài thi.

## Mất cân bằng giữa các khối thi

### Số lượng thí sinh chọn khối thi

**15 điểm** là mức điểm trung bình của phần lớn các khối thi nên đây sẽ là mức “điểm sàn” để so sánh điểm giữa các khối thi trong xét tuyển CĐ, ĐH.

Có **191,593** thí sinh chọn khối **C00** (chiếm 35%), theo sau là khối **D01** có **148,896** thí sinh (chiếm 27%), hai khối thi tự nhiên là **A00** và **B00** có lần lượt số thí sinh là **111,019** (chiếm 20%) và **59,856** (chiếm 11%), cuối cùng là khối **A01** có **39,240** thí sinh (chiếm 7%).



Số lượng thí sinh của mỗi khối là khối có điểm xét thi cao nhất  
(Thí sinh có tổng điểm thi  $\geq 15$ )

%	A00	A01	B00	C00	D01
1%	24.5	25	24.5	24.25	23.75
5%	22.75	23.75	22.5	21.75	22.75
10%	21.75	22.75	21.5	20.75	21.25
15%	21	22.25	20.75	19.5	20.5
20%	20.25	21.5	20	19	20
30%	19.25	20.5	19	18.25	19

Bảng phân bố điểm thi của các khối thi  
(Thí sinh có tổng điểm thi  $\geq 15$ )

### So sánh nhóm điểm cao giữa các khối thi

**Khối A01** có số lượng thí sinh thấp nhất nhưng khi xét ở nhóm điểm cao, điểm thi ở các nhóm đều cao vượt trội so với các khối thi còn lại, đặc biệt khi so với 2 khối thi gần (khác biệt 1 môn thi) như A00 và D01.

**Khối C00** do có ảnh hưởng từ môn **Lịch Sử** nên điểm thi ở các nhóm thấp hơn các nhóm còn lại 0.5 - 1 điểm.

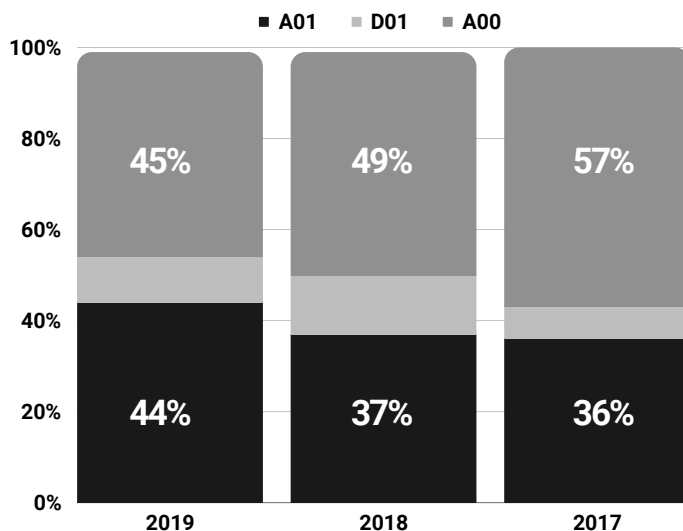
Các khối còn lại nhìn chung không có chênh lệch lớn.

### Chênh lệch điểm ảnh hưởng đến xét tuyển Đại Học

Theo dữ liệu điểm thi trúng tuyển theo các khối ở trường ĐH-CNTT, với đặc thù là nhóm ngành kỹ thuật nên phần lớn thí sinh sẽ chọn khối **A00**.

Tuy nhiên với quy chế tuyển sinh thì điểm thi giữa các khối thi là như nhau, điều này sẽ làm cho nhóm thí sinh khối **A00** gặp bất lợi khi cạnh tranh với khối **A01** ở các nhóm ngành có điểm chuẩn cao [Phụ lục 1-2].

Xét 3 năm thi gần có hình thức thi như sau thì tỷ lệ thí sinh trúng tuyển bằng khối **A01** chiếm cao hơn qua mọi năm mặc dù số lượng thí sinh thi khối **A01** là rất thấp.



Tỷ lệ thí sinh trúng tuyển vào trường ĐH CNTT - ĐHQG TP.HCM theo các khối A00, A01, D01 từ năm 2017 - 2019

## 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 tăng tỷ lệ thí sinh đạt trung bình cao thì sẽ vô tình tịnh tiến phổ điểm sang phải làm số lượng điểm giỏi, xuất sắc tăng cao.

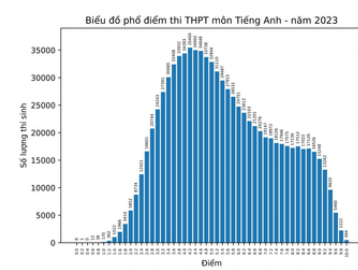
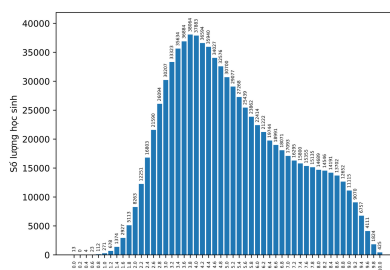
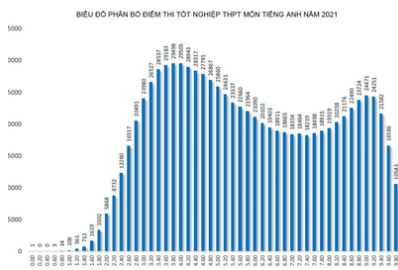
	Toán	Anh	Văn	Hóa	Lý	Sinh	Sử	Địa	GD&CD
<b>2017</b>	281	996	1	1521	75	401	107	603	250
<b>2018</b>	2	106	0	15	3	2	10	32	309

*Số lượng điểm 10 của 2 năm liên tiếp 2017, 2018 có sự chênh lệch lớn*

## Thi 6 môn là quá nhiều dẫn đến phổ điểm không chính xác

Như đã đề cập ở đầu bài viết, 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).

Điều này góp phần làm lệch kết quả phản ánh chính xác của đề thi và khó để điều chỉnh, ảnh hưởng lớn nhất là môn **tiếng Anh**, tỷ lệ thí sinh chọn **tiếng Anh** là môn xét tuyển Đại Học thực tế chỉ chiếm **33%** thí sinh dự thi. Việc điều chỉnh đề thi dễ hơn có thể dẫn đến tình trạng **mưa điểm 10** hay tệ hơn là **hình thành một phổ điểm có 2 đỉnh ở nhóm điểm thấp và nhóm điểm cao** thay vì tịnh tiến điểm trung vị như những năm tổ chức kỳ thi tốt nghiệp THPT vài năm gần đây, khiến chúng ta lầm tưởng về chất lượng học của học sinh nhưng thực tế việc chênh lệch học lực đã có từ nhiều năm trước nhưng mức độ hiện diện không rõ ràng.



*Phổ điểm môn tiếng Anh kỳ thi tốt nghiệp THPT 3 năm gần nhất, cạnh tranh ở từng câu trắc nghiệm*

## Cần đảm bảo cân bằng độ khó giữa các môn thi

Đối với xét tuyển đại học, việc các môn thi có độ khó như nhau là cần thiết vì tùy theo nhóm ngành mà sẽ có một số điều kiện khác nhau như nhân hai điểm môn chính, xét điểm các khối như sau. Nếu tỉ lệ thí sinh đạt mức điểm của các môn có chênh lệch cao sẽ ảnh hưởng đến quyền lợi của nhiều thí sinh.



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>

Tuấn T.M. (2012). ***Giáo trình Thiết kế Cơ sở dữ liệu***. NXB ĐHQG TP HCM.  
<https://ir.vnulib.edu.vn/handle/VNUHCM/6665>

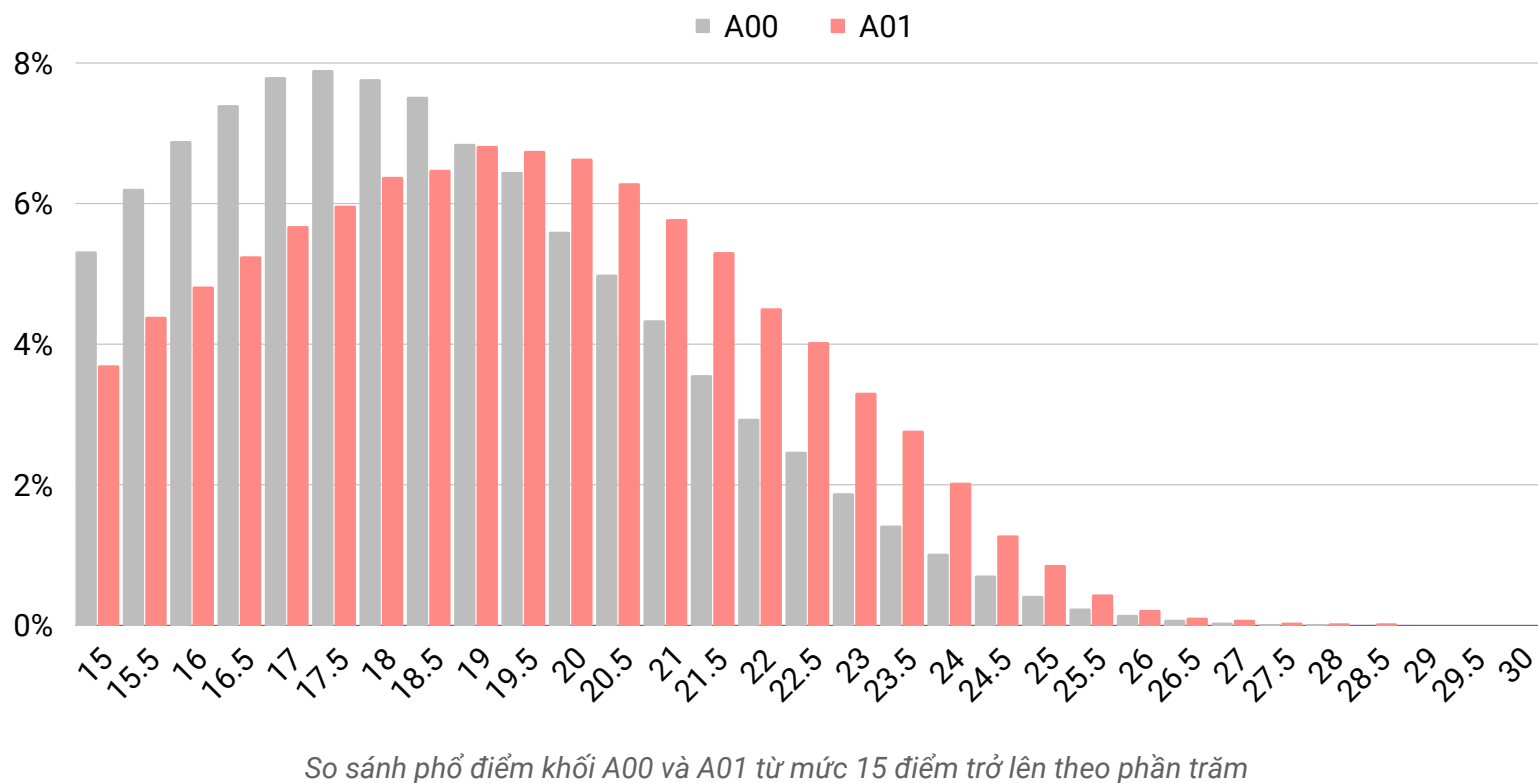
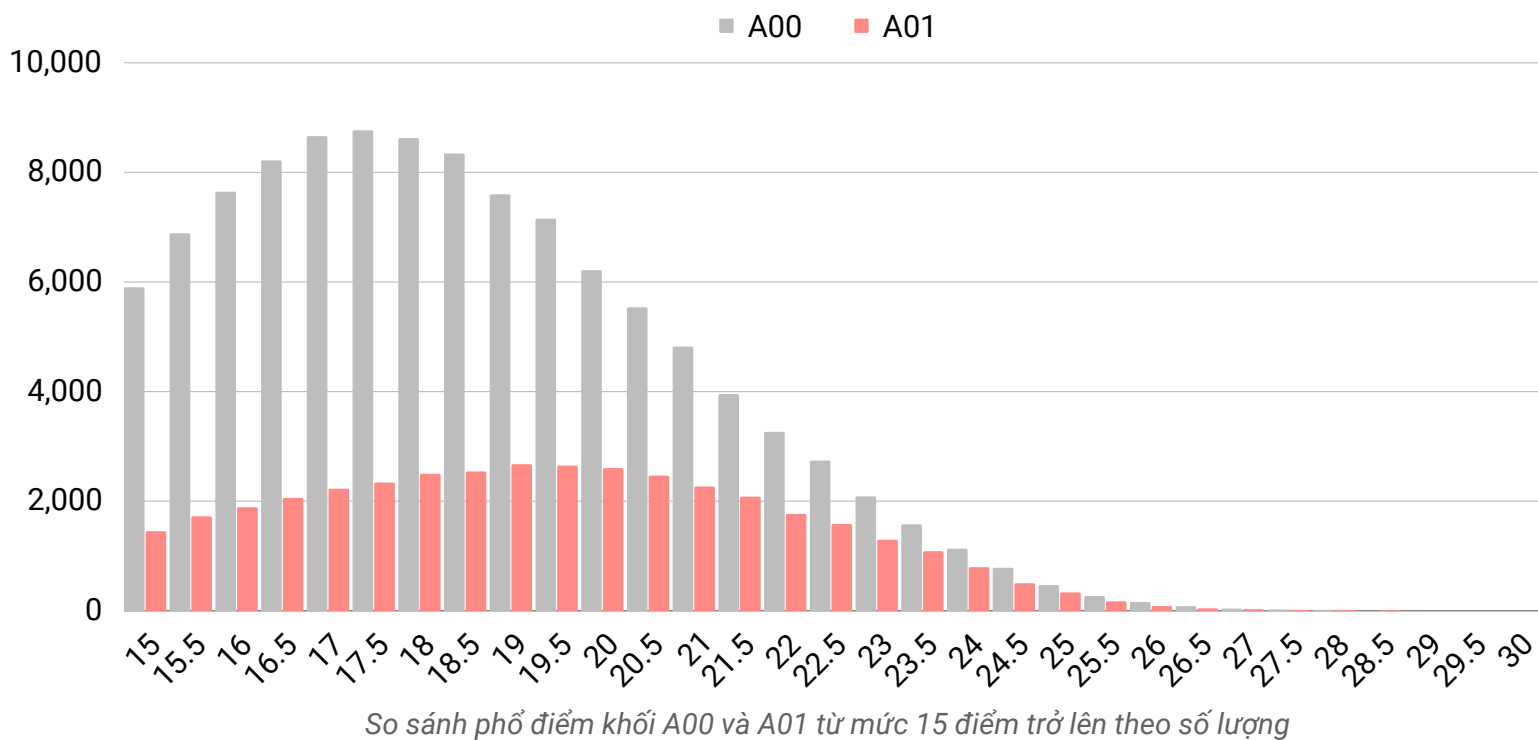
Trần, D. P., & Vũ, C. T. (2023). ***Tiêu chuẩn định lượng cho một bảng điểm kết thúc học phần đẹp***.  
Tạp Chí Nghiên cứu Khoa học Và Phát triển, 2(1), 35-40.  
<https://doi.org/10.58902/tcnckhpt.v2i1.33>

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/>

***Hà Giang: Hơn 330 bài thi có tổng điểm chênh lệch so với điểm chấm (2018)***. Retrieved Aug 23,  
2023, from <https://hoabinhtv.vn/van-hoa-xa-hoi/ha-giang-hon-330-bai-thi-co-tong-diem-chenh-lech-so-voi-diem-cham>

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-nghiep-thpt-quoc-gia-2018-cao-nhat-trong-4-nam-gan-day-13027.html>

## Phụ lục 1: So sánh chi tiết phổ điểm thực tế của khối A00 và khối A01



## Mức độ ảnh hưởng của vụ việc gian lận điểm thi lên phổ điểm

**Các bài thi được chỉnh sửa đều ở mức trên 26 điểm trở lên và chiếm tỷ lệ thấp nên nhìn chung không có ảnh hưởng đến bài phân tích này.**

**Phụ lục 2: Điểm trúng tuyển trường ĐH Công nghệ Thông tin - ĐHQG TP.HCM năm 2018**

STT	Ngành	Điểm chuẩn	Số lượng thí sinh trúng tuyển			Tỷ lệ thí sinh trúng tuyển		
			A00	A01	D01	A00	A01	D01
1	Kỹ thuật phần mềm	23.2	30	24	1	55%	44%	2%
2	Công nghệ thông tin	22.5	64	18	16	65%	18%	16%
3	An toàn thông tin	22.5	29	20	3	56%	38%	6%
4	Khoa học máy tính	22.4	37	20	4	61%	33%	7%
5	Kỹ thuật máy tính	21.7	39	25	2	59%	38%	3%
6	Kỹ thuật phần mềm (CLC)	21.5	31	31	4	47%	47%	6%
7	Thương mại điện tử	21.2	26	17	20	41%	27%	32%
8	Hệ thống thông tin	21.1	28	13	5	61%	28%	11%
9	Công nghệ thông tin (Khoa học dữ liệu)	20.6	29	14	7	58%	28%	14%
10	Khoa học máy tính (CLC)	20.25	45	54	6	43%	51%	6%
12	An toàn thông tin (CLC)	20.1	23	30	8	38%	49%	13%
13	Công nghệ thông tin (CLC Nhật Bản)	19	27	37	9	37%	51%	12%
14	Hệ thống thông tin (CLC)	19	31	21	11	49%	33%	17%
15	Mạng máy tính và TTDL (CLC)	18.6	29	23	20	40%	32%	28%
16	Kỹ thuật máy tính (CLC)	18.4	51	49	11	46%	44%	10%
17	Hệ thống thông tin (TT)	17	22	12	17	43%	24%	33%

**Các thông tin cần thiết khác để thí sinh ĐKXT vào các ngành của trường: mã số trường, mã số ngành, tổ hợp xét tuyển và quy định chèn lịch điểm xét tuyển giữa các tổ hợp; các điều kiện phụ sử dụng trong xét tuyển...**

- Điểm xét tuyển là tổng điểm 03 môn thi của tổ hợp môn xét tuyển (không nhân hệ số) cộng điểm ưu tiên khu vực và đối tượng.
- Điểm trúng tuyển cho các tổ hợp môn xét tuyển khác nhau của cùng 01 ngành và cùng chương trình xét tuyển là như nhau.