

# Mục lục

<b>1</b>	<b>Mở đầu</b>	<b>2</b>
1.1	Bối cảnh và ý nghĩa đề tài . . . . .	2
1.2	Mục tiêu nghiên cứu . . . . .	2
1.3	Phạm vi và đối tượng dữ liệu . . . . .	2
<b>2</b>	<b>Dữ liệu và công cụ</b>	<b>3</b>
2.1	Nguồn dữ liệu . . . . .	3
2.2	Cấu trúc dữ liệu . . . . .	3
2.3	Công cụ và thư viện sử dụng . . . . .	3
2.4	Khôi phục và chuẩn hoá mã SBD . . . . .	3
2.5	Xử lý giá trị thiếu . . . . .	4
2.6	Tỉ lệ thí sinh đạt điểm 0 và tỉ lệ bỏ thi . . . . .	4
2.7	Chuẩn hoá kiểu dữ liệu và phát hiện giá trị bất thường . . . . .	6
<b>3</b>	<b>Thống kê mô tả cho từng môn</b>	<b>6</b>
3.1	Các chỉ tiêu trung tâm . . . . .	6
3.2	Các chỉ tiêu phân tán: min, max, độ lệch chuẩn, phương sai và khoảng biến thiên . . . . .	8
3.3	Tỷ lệ đạt các mức điểm . . . . .	10
<b>4</b>	<b>Phân phối điểm</b>	<b>19</b>
4.1	Histogram các môn . . . . .	19
4.2	Theo tỉnh và vùng . . . . .	27
4.3	Giới thiệu các tổ hợp xét tuyển . . . . .	37
4.4	Biểu đồ phân phối điểm theo tổ hợp . . . . .	37
<b>5</b>	<b>Kiểm tra tương quan giữa các môn</b>	<b>39</b>
5.1	Ma trận tương quan . . . . .	39
<b>6</b>	<b>Những nhận định chính</b>	<b>41</b>
6.1	Phân phối điểm và mức độ phân hoá . . . . .	41
6.2	Độ biến động và rủi ro theo môn . . . . .	41
6.3	Tổ hợp xét tuyển và đặc trưng phân phối điểm . . . . .	42
6.4	Khác biệt theo tỉnh và vùng . . . . .	42
6.5	Tương quan giữa các nhóm môn . . . . .	42
6.6	Tỷ lệ đậu – rớt tốt nghiệp . . . . .	42
<b>7</b>	<b>Đề xuất và hướng cải thiện điểm</b>	<b>42</b>
7.1	Tăng cường ôn luyện các môn có rủi ro cao . . . . .	42
7.2	Lựa chọn tổ hợp xét tuyển theo điểm mạnh . . . . .	43
7.3	Ôn tập theo nhóm môn có tương quan . . . . .	43
7.4	Giảm thiểu nguy cơ điểm liệt . . . . .	43
7.5	Chiến lược nâng điểm theo mốc mục tiêu . . . . .	43
<b>8</b>	<b>Tổng kết</b>	<b>43</b>

# 1 Mở đầu

## 1.1 Bối cảnh và ý nghĩa đề tài

Kỳ thi THPT quốc gia năm 2018 là một trong những kỳ thi quan trọng nhất trong hệ thống giáo dục Việt Nam, vừa dùng để xét tốt nghiệp THPT, vừa là căn cứ tuyển sinh đại học, cao đẳng. Kết quả điểm thi phản ánh:

- Mức độ đạt được mục tiêu chương trình phổ thông.
- Sự khác biệt về chất lượng giáo dục giữa các vùng miền, tỉnh/thành.
- Ảnh hưởng của tổ hợp môn xét tuyển, xu hướng học tập các khối thi truyền thống và mới.

Việc xử lý, phân tích dữ liệu điểm thi một cách hệ thống giúp:

- Hiểu rõ hơn phân phối điểm theo từng môn.
- Đánh giá mức độ khó/dễ của đề thi.
- Phát hiện các chênh lệch vùng miền, nhóm môn học.
- Gợi ý chính sách giáo dục, ôn luyện thi hiệu quả hơn cho các năm sau.

## 1.2 Mục tiêu nghiên cứu

Đề tài hướng tới các mục tiêu chính:

- Xây dựng quy trình chuẩn để làm sạch và xử lý dữ liệu điểm thi THPT.
- Thực hiện thống kê mô tả cho từng môn học: điểm trung bình, độ phân tán, tỉ lệ đạt các mức điểm.
- Phân tích phân phối điểm theo từng môn và toàn bộ hệ thống môn thi.
- So sánh kết quả thi theo tỉnh/thành phố, vùng miền và theo tổ hợp xét tuyển.
- Tính tổng điểm xét tuyển, xếp hạng và phân tích sự khác biệt giữa các nhóm.
- Kiểm tra tương quan giữa các môn thi nhằm tìm hiểu mối quan hệ giữa các kỹ năng, lĩnh vực kiến thức.
- Xây dựng báo cáo tổng hợp kèm bảng biểu, biểu đồ và phụ lục mã nguồn.

## 1.3 Phạm vi và đối tượng dữ liệu

**Dữ liệu:** Điểm thi THPT quốc gia năm 2018.

**Phạm vi nghiên cứu** được xác định linh hoạt tùy mục tiêu phân tích, có thể bao quát:

- Toàn bộ thí sinh trên phạm vi cả nước

**Đối tượng nghiên cứu** là dữ liệu điểm thi của thí sinh tham gia Kỳ thi THPT Quốc gia, bao gồm các môn: Toán, Ngữ văn, Ngoại ngữ, Vật lý, Hóa học, Sinh học, Lịch sử, Địa lý và Giáo dục công dân.

## 2 Dữ liệu và công cụ

### 2.1 Nguồn dữ liệu

Dữ liệu phục vụ nghiên cứu được thu thập từ:

- Các tệp điểm thi do Bộ Giáo dục và Đào tạo công bố.
- Những nguồn dữ liệu thứ cấp như báo cáo thống kê, kho dữ liệu mở và các nguồn công khai khác.

Mỗi dòng dữ liệu đại diện cho một thí sinh, bao gồm các thông tin:

- Số báo danh (SBD).
- Mã tỉnh hoặc mã điểm thi.
- Điểm số của từng môn thi.
- Thông tin tổ hợp xét tuyển (nếu có).

### 2.2 Cấu trúc dữ liệu

Các trường dữ liệu chính thường bao gồm:

- SBD: chuỗi ký tự độ dài 8–9 ký tự.
- Các cột điểm: TOAN, VAN, ANH, LY, HOA, SINH, SU, DIA, GD CD.

### 2.3 Công cụ và thư viện sử dụng

Ngôn ngữ: Python.

Thư viện:

- pandas dùng để xử lý, làm sạch và thao tác dữ liệu dạng bảng.
- numpy hỗ trợ các phép toán số học và mảng.
- matplotlib, seaborn, plotly phục vụ trực quan hóa dữ liệu.
- scipy hỗ trợ phân tích thống kê nâng cao.

### 2.4 Khôi phục và chuẩn hoá mã SBD

Số báo danh (SBD) đóng vai trò là biến nhận dạng duy nhất của mỗi thí sinh trong bộ dữ liệu. Các thao tác phân nhóm theo tỉnh, liên kết dữ liệu, hoặc kiểm tra tính nhất quán đều dựa vào SBD. Vì vậy, việc chuẩn hoá và khôi phục SBD là bước nền tảng nhằm đảm bảo độ chính xác của toàn bộ quy trình phân tích. Quy trình thực hiện bao gồm:

- **Chuẩn hoá SBD về dạng chuỗi (string).** Trong một số bộ dữ liệu, SBD có thể bị hệ thống đọc sai thành số nguyên, dẫn đến mất số 0 ở đầu. Việc chuyển toàn bộ SBD về kiểu chuỗi giúp bảo toàn đầy đủ ký tự và tránh lỗi khi xử lý.

- **Bổ sung số 0 ở đầu để đồng nhất độ dài.** Các SBD không đủ 8–9 ký tự được bổ sung số 0 ở đầu nhằm chuẩn hoá định dạng. Điều này đảm bảo tính nhất quán khi tách mã tỉnh và hạn chế các sai lệch liên quan đến căn chỉnh ký tự.
- **Tách mã tỉnh từ SBD theo quy ước.** Hai ký tự đầu của SBD phản ánh mã tỉnh (01–63). Từ đó có thể xây dựng biến MA\_TINH để phục vụ phân tích theo địa phương, so sánh chất lượng giáo dục giữa các tỉnh/thành phố.
- **Kiểm tra và xử lý các SBD trùng lặp hoặc sai định dạng.** Một số SBD có thể bị nhập trùng hoặc chứa ký tự không hợp lệ. Các trường hợp này được rà soát và xử lý dựa trên quy tắc: loại bỏ bản ghi, hiệu chỉnh theo mẫu chuẩn hoặc gắn nhãn dữ liệu lỗi để xử lý riêng. Việc kiểm tra này giúp duy trì tính toàn vẹn của dữ liệu trước khi chuyển sang các bước phân tích tiếp theo.

## 2.5 Xử lý giá trị thiếu

Trong bất kỳ bộ dữ liệu thực tế nào, đặc biệt là dữ liệu điểm thi, sự xuất hiện của giá trị thiếu (NA) là điều thường gặp. Việc xử lý chính xác các giá trị này giúp đảm bảo tính tin cậy của kết quả phân tích. Các bước xử lý bao gồm:

- **Đếm số lượng giá trị thiếu theo từng môn.** Với mỗi cột điểm, thống kê số lượng NA để đánh giá mức độ ảnh hưởng của thiếu dữ liệu.
- **Tính tỉ lệ phần trăm giá trị thiếu.** Việc tính tỉ lệ NA giúp so sánh mức độ thiếu giữa các môn có quy mô dữ liệu khác nhau. Ví dụ, các môn tự chọn hoặc môn Ngoại ngữ đôi khi có số lượng NA cao hơn.
- **Lựa chọn chiến lược xử lý phù hợp.** Tùy mục tiêu phân tích, có thể:
  - Loại bỏ các bản ghi thiếu toàn bộ điểm.
  - Giữ lại nhưng gán NA và loại bỏ khỏi các phép tính thống kê khi cần.

## 2.6 Tỉ lệ thí sinh đạt điểm 0 và tỉ lệ bỏ thi

Để đánh giá chất lượng dữ liệu và mức độ tham gia thi của thí sinh ở từng môn, hai chỉ tiêu quan trọng được tính toán gồm: tỉ lệ đạt điểm 0 và tỉ lệ bỏ thi (giá trị NA). Các chỉ tiêu này giúp nhận diện mức độ khó của đề thi, mức độ lựa chọn môn tự chọn và chất lượng của dữ liệu gốc.

**Tỉ lệ thí sinh đạt điểm 0.** Tỉ lệ điểm 0 của từng môn được xác định bằng cách đếm số bài thi có điểm bằng 0 và chia cho tổng số thí sinh dự thi môn đó. Kết quả cho thấy phần lớn các môn có tỉ lệ điểm 0 rất thấp, thường chỉ từ 0.0001% đến 0.0006%. Điều này phản ánh rằng thí sinh hiếm khi bị điểm tuyệt đối 0, trừ trường hợp bỏ trống bài làm hoặc có lỗi phát sinh khi nhập dữ liệu.

**Tỉ lệ bỏ thi.** Tỉ lệ bỏ thi được tính bằng tỉ lệ giá trị thiếu (NA) trên tổng số thí sinh. Các môn không bắt buộc, đặc biệt là khối các môn tự chọn, có tỉ lệ NA rất cao do không phải thí sinh nào cũng đăng ký dự thi. Nhiều môn có tỉ lệ bỏ thi trên 70%, thậm chí trên 90%. Ngược lại, các môn bắt buộc như Toán và Ngữ văn có tỉ lệ NA rất thấp, dưới 0.5%.

```
[28]: ty_le_diem_0 = (df_all[subjects] == 0).mean() * 100
      ty_le_diem_0
```

```
[28]: Toán                0.0004
      Văn                0.0006
      Lí                 0.0001
      Hóa                0.0000
      Sinh              0.0000
      Tin học           0.0000
      Công nghệ công nghiệp 0.0000
      Công nghệ nông nghiệp 0.0000
      Sử                0.0002
      Địa               0.0003
      Giáo dục kinh tế và pháp luật 0.0000
      Ngoại ngữ         0.0002
      dtype: float64
```

```
[29]: ty_le_bo_thi = df_all[subjects].isna().mean() * 100
      ty_le_bo_thi
```

```
[29]: Toán                0.4301
      Văn                0.3858
      Lí                 69.0118
      Hóa                79.0448
      Sinh              94.3270
      Tin học           99.3943
      Công nghệ công nghiệp 99.8061
      Công nghệ nông nghiệp 98.1769
      Sử                57.8396
      Địa               57.8996
      Giáo dục kinh tế và pháp luật 78.5426
      Ngoại ngữ         67.0122
      dtype: float64
```

```
[30]: def central_stats(series):
      series = series.dropna()
      return {
          "mean": series.mean(),
          "median": series.median(),
          "mode": series.mode().iloc[0] if not series.mode().empty else None
      }
```

```
[31]: stats = {}

      for col in subjects:
          stats[col] = central_stats(df_all[col])
```

Hình 1: Kết quả thống kê tỉ lệ điểm 0 và tỉ lệ bỏ thi.

## 2.7 Chuẩn hoá kiểu dữ liệu và phát hiện giá trị bất thường

Sau khi xử lý các giá trị thiếu, bước tiếp theo là chuẩn hoá kiểu dữ liệu của các cột điểm thi nhằm đảm bảo tính chính xác trong toàn bộ quá trình phân tích. Các cột điểm có thể chứa ký tự không hợp lệ, hoặc bị hệ thống nhận dạng sai kiểu dữ liệu. Việc chuẩn hoá giúp loại bỏ những sai lệch phát sinh trong quá trình nhập liệu hoặc tổng hợp dữ liệu.

- **Chuẩn hoá dữ liệu về dạng số thực.** Tất cả các cột điểm được ép kiểu về float. Những giá trị không hợp lệ như chuỗi rỗng, dấu chấm hoặc ký hiệu đặc biệt được chuyển thành giá trị thiếu (NA) để thuận tiện cho các phép tính thống kê.
- **Phát hiện và loại bỏ các điểm ngoài phạm vi quy định.** Theo quy định, điểm thi nằm trong khoảng từ 0 đến 10. Các điểm âm hoặc điểm lớn hơn 10 được xem là bất thường và được chuyển lại thành NA. Một số địa phương sử dụng giá trị -1 để đánh dấu thí sinh không dự thi; trường hợp này cũng được chuyển đổi về giá trị thiếu.
- **Kiểm tra tính đồng nhất giữa các môn.** Việc kiểm tra bao gồm đánh giá số chữ số thập phân, các bước nhảy điểm lẻ (đặc biệt đối với các môn trắc nghiệm thường có bước 0.25), và nhận diện các điểm bất thường không tuân theo quy tắc chấm thi.
- **Đánh giá chất lượng dữ liệu sau chuẩn hoá.** Thống kê tổng số lượng điểm bị loại hoặc chuyển thành NA giúp xác định mức độ ảnh hưởng của các lỗi nhập liệu đến từng môn và toàn bộ dữ liệu. Đây là cơ sở để kiểm tra độ tin cậy của tập dữ liệu trước khi thực hiện các phân tích tiếp theo.

## 3 Thống kê mô tả cho từng môn

### 3.1 Các chỉ tiêu trung tâm

Các chỉ tiêu trung tâm (mean, median, mode) cho phép mô tả xu hướng tập trung của phân phối điểm thi ở từng môn. Dựa trên kết quả tính toán, có thể rút ra nhiều nhận xét quan trọng về mức độ khó dễ và mức độ phân hoá của bài thi.

**Môn Toán.** Điểm trung bình đạt 4.84, thấp hơn trung vị 4.75 và mode 4.75, cho thấy phân phối hơi lệch phải và đa số thí sinh gặp khó khăn ở môn này. Môn Toán có mức độ phân hoá cao và là môn gây áp lực lớn nhất.

**Môn Ngữ văn.** Điểm trung bình 7.00 và trung vị 7.25 cho thấy mặt bằng điểm khá cao. Đây là môn tự luận nên mức độ phân tán thấp và kết quả ổn định hơn.

**Môn Vật lý.** Mean  $\approx$  median  $\approx$  mode cho thấy phân phối cân bằng. Điểm cao xuất hiện nhiều, phản ánh đặc thù trắc nghiệm rõ ràng và phương pháp ôn luyện hiệu quả.

**Môn Hoá học.** Kết quả ổn định, không quá cao cũng không quá thấp. Phân phối tương đối đều.

**Môn Sinh học.** Phân phối hơi lệch phải, nhiều điểm trung bình, ít điểm cao. Thể hiện mức độ khó tương đối.

**Môn Tin học và nhóm Công nghệ.** Điểm trung bình cao, thí sinh đạt điểm khá–giỏi nhiều. Đây thường là nhóm thí sinh có năng lực tốt.

**Sử, Địa, GDCD.** GDCD nổi bật với điểm trung bình cao nhất trong nhóm xã hội. Phân phối nghiêng về điểm giỏi.

**Ngoại ngữ.** Điểm trung bình thấp nhất trong toàn bộ các môn ( $\approx 5.43$ ). Điều này phản ánh sự phân hoá mạnh giữa nhóm thí sinh có nền tảng ngoại ngữ tốt và nhóm yếu.

```
df_center = pd.DataFrame(stats).T
df_center
```

[31]:

	mean	median	mode
Toán	4.836976	4.75	4.75
Văn	7.009409	7.25	7.50
Lí	7.024686	7.00	7.50
Hóa	6.117026	6.00	6.25
Sinh	5.810411	5.75	6.50
Tin học	6.824847	6.85	7.00
Công nghệ công nghiệp	5.678804	5.50	4.75
Công nghệ nông nghiệp	7.674659	7.75	8.00
Sử	6.524415	6.60	7.75
Địa	6.629437	6.75	7.75
Giáo dục kinh tế và pháp luật	7.696897	7.75	8.25
Ngoại ngữ	5.439699	5.25	5.25

```
[32]: def spread_stats(series):
        s = series.dropna()

        return {
            "min": s.min(),
            "max": s.max(),
            "std": s.std(),          # độ lệch chuẩn
            "var": s.var(),          # phương sai
            "range": s.max() - s.min()
        }
```

[33]:

```
spread = {}

for col in subjects:
    spread[col] = spread_stats(df_all[col])

df_spread = pd.DataFrame(spread).T
df_spread
```

[33]:

	min	max	std	var	range
Toán	0.00	10.00	1.696457	2.877966	10.00
Văn	0.00	9.75	1.294472	1.675659	9.75
Lí	0.00	10.00	1.518879	2.306993	10.00
Hóa	0.75	10.00	1.815658	3.296612	9.25
Sinh	0.85	10.00	1.613341	2.602870	9.15
Tin học	1.35	10.00	1.495923	2.237786	8.65
Công nghệ công nghiệp	2.30	10.00	1.513286	2.290036	7.70
Công nghệ nông nghiệp	1.55	10.00	1.188451	1.412416	8.45
Sử	0.00	10.00	1.643521	2.701161	10.00
Địa	0.00	10.00	1.762820	3.107534	10.00

8

Hình 4.1. Bảng giá trị trung bình (mean), trung vị (median) và một (mode) của từng môn thi.

### 3.2 Các chỉ tiêu phân tán: min, max, độ lệch chuẩn, phương sai và khoảng biến thiên

Bên cạnh các chỉ tiêu trung tâm, việc phân tích các chỉ tiêu phân tán cho phép đánh giá độ rộng và mức độ lan toả của phân phối điểm thi ở từng môn. Các chỉ tiêu này gồm giá trị nhỏ nhất, giá trị lớn nhất, độ lệch chuẩn, phương sai và khoảng biến thiên.

**Giá trị nhỏ nhất và lớn nhất.** Hầu hết các môn có giá trị nhỏ nhất bằng 0, phản ánh việc vẫn có thí sinh bỏ bài hoặc không hoàn thành bài thi. Các môn trắc nghiệm như Lý, Hóa, Sinh và Ngoại ngữ có giá trị lớn nhất bằng 10, cho phép đạt mức điểm tuyệt đối. Riêng môn Ngữ văn có giá trị lớn nhất là 9.75, phù hợp với đặc thù chấm tự luận khó đạt điểm tuyệt đối.

**Độ lệch chuẩn.** Độ lệch chuẩn cho biết mức độ dao động của điểm thi. Các môn Toán, Hóa, Sinh và Địa có độ lệch chuẩn cao, cho thấy sự phân hoá mạnh giữa các nhóm thí sinh. Ngược lại, môn Ngữ văn và một số môn tự chọn có độ lệch chuẩn thấp, phản ánh sự ổn định và ít biến động trong phân phối điểm.

**Phương sai.** Phương sai cao ở Toán, Hóa và Địa tiếp tục khẳng định mức độ phân tán lớn của các môn này. Ngữ văn và một số môn công nghệ có phương sai thấp, thể hiện phân phối điểm tập trung hơn.

**Khoảng biến thiên.** Hầu hết các môn có khoảng biến thiên bằng 10, do điểm trải dài từ 0 đến 10. Tuy nhiên, khoảng biến thiên lớn không phản ánh mức phân tán thực tế nếu độ lệch chuẩn thấp; do đó cần kết hợp các chỉ tiêu để đánh giá đầy đủ.

```
df_center = pd.DataFrame(stats).T
df_center
```

```
[31]:
```

	mean	median	mode
Toán	4.836976	4.75	4.75
Văn	7.009409	7.25	7.50
Lí	7.024686	7.00	7.50
Hóa	6.117026	6.00	6.25
Sinh	5.810411	5.75	6.50
Tin học	6.824847	6.85	7.00
Công nghệ công nghiệp	5.678804	5.50	4.75
Công nghệ nông nghiệp	7.674659	7.75	8.00
Sử	6.524415	6.60	7.75
Địa	6.629437	6.75	7.75
Giáo dục kinh tế và pháp luật	7.696897	7.75	8.25
Ngoại ngữ	5.439699	5.25	5.25

```
[32]: def spread_stats(series):
    s = series.dropna()

    return {
        "min": s.min(),
        "max": s.max(),
        "std": s.std(),           # độ lệch chuẩn
        "var": s.var(),          # phương sai
        "range": s.max() - s.min()
    }
```

```
[33]: spread = {}

for col in subjects:
    spread[col] = spread_stats(df_all[col])

df_spread = pd.DataFrame(spread).T
df_spread
```

```
[33]:
```

	min	max	std	var	range
Toán	0.00	10.00	1.696457	2.877966	10.00
Văn	0.00	9.75	1.294472	1.675659	9.75
Lí	0.00	10.00	1.518879	2.306993	10.00
Hóa	0.75	10.00	1.815658	3.296612	9.25
Sinh	0.85	10.00	1.613341	2.602870	9.15
Tin học	1.35	10.00	1.495923	2.237786	8.65
Công nghệ công nghiệp	2.30	10.00	1.513286	2.290036	7.70
Công nghệ nông nghiệp	1.55	10.00	1.188451	1.412416	8.45
Sử	0.00	10.00	1.643521	2.701161	10.00
Địa	0.00	10.00	1.762820	3.107534	10.00

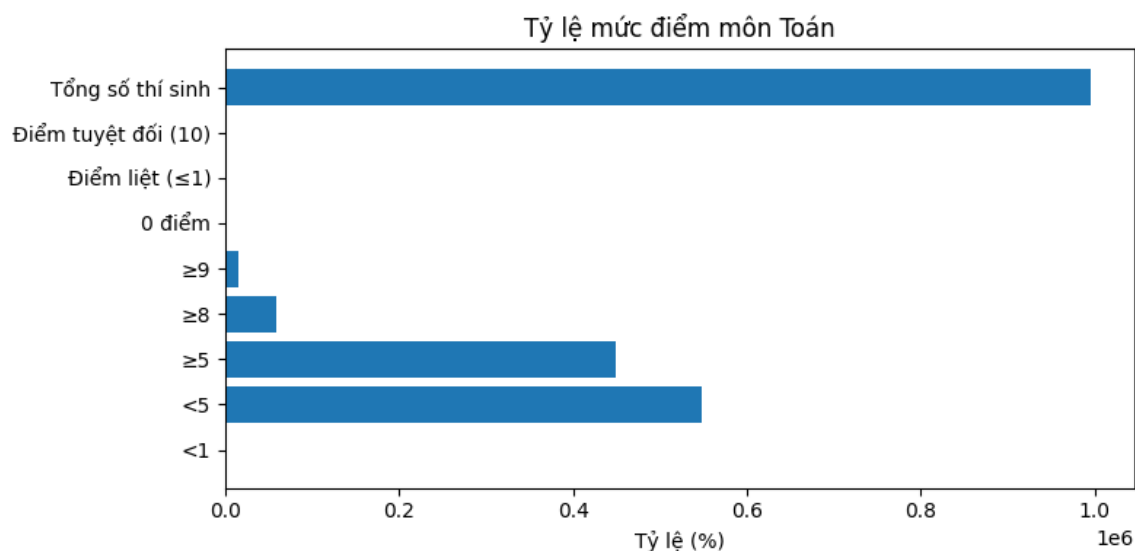
8

Thống kê min, max, độ lệch chuẩn, phương sai và khoảng biến thiên theo từng môn.

### 3.3 Tỷ lệ đạt các mức điểm

Việc phân tích tỷ lệ thí sinh đạt theo các mức điểm cho phép đánh giá mức độ khó của đề thi, khả năng phân hoá của từng môn và chất lượng làm bài của thí sinh. Các mức điểm được xem xét gồm: dưới 1 điểm, dưới 5 điểm, từ 5 trở lên, từ 8 trở lên, từ 9 trở lên, điểm

0, điểm liệt và điểm tuyệt đối 10. Các ngưỡng này phản ánh đầy đủ các nhóm năng lực từ rất yếu đến xuất sắc.



Tỷ lệ đạt các mức điểm môn Toán.

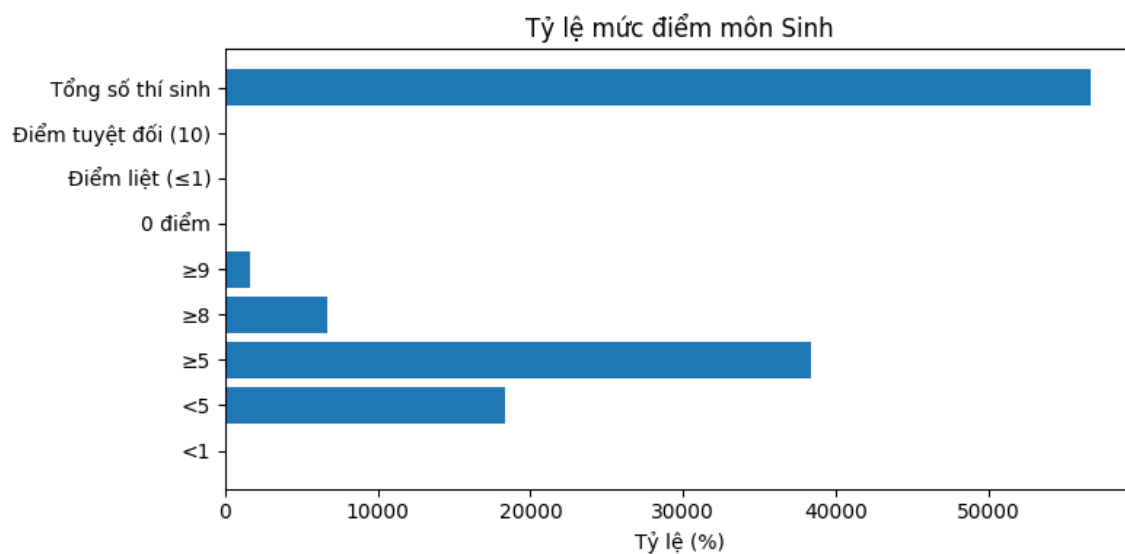
## Môn Toán

**Nhận xét.** Biểu đồ cho thấy mức độ phân hoá rất mạnh của môn Toán. Khoảng 55% thí sinh đạt dưới 5 điểm, cho thấy đề thi khó và thí sinh gặp nhiều hạn chế ở môn này. Tỷ lệ điểm dưới 1 cũng đáng chú ý, phản ánh một nhóm thí sinh có năng lực yếu hoặc mất căn bản.

Trong khi đó, khoảng 45% thí sinh đạt từ 5 điểm trở lên, chủ yếu tập trung ở mức trung bình và khá. Nhóm điểm cao ( $\geq 8$  và  $\geq 9$ ) chiếm tỷ lệ rất nhỏ, cho thấy số thí sinh đạt mức giỏi trong môn Toán không nhiều. Tỷ lệ điểm tuyệt đối 10 gần như bằng 0, thể hiện đề thi năm 2018 có mức độ phân loại cao.

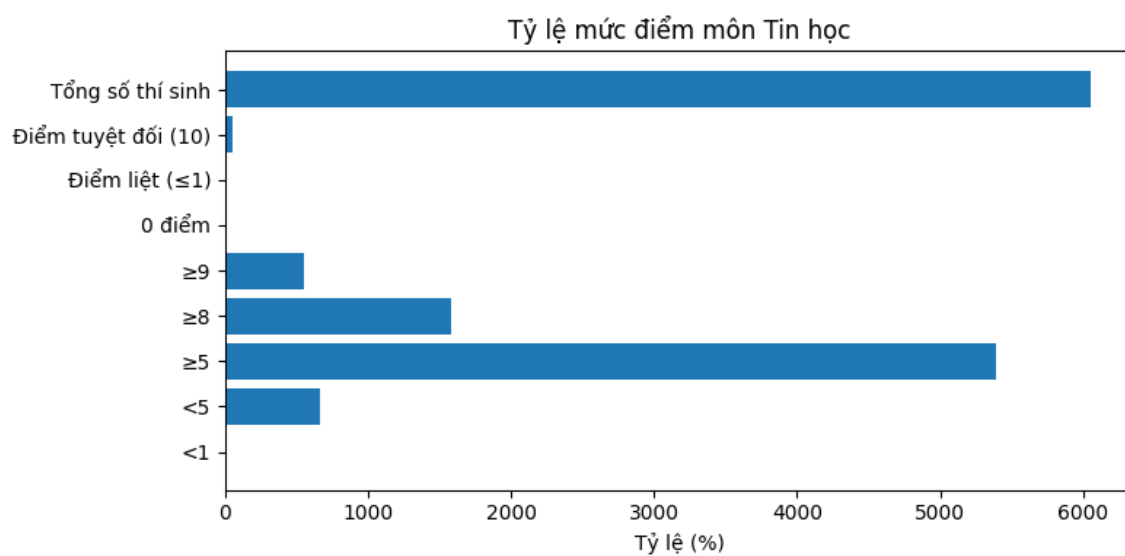
Điểm 0 và điểm liệt xuất hiện nhưng không nhiều, tuy nhiên vẫn phản ánh một phần khó khăn chung của môn thi này. Đây là môn có độ khó lớn nhất trong kỳ thi và có vai trò quyết định trong năng lực logic của thí sinh.

## Môn Sinh học



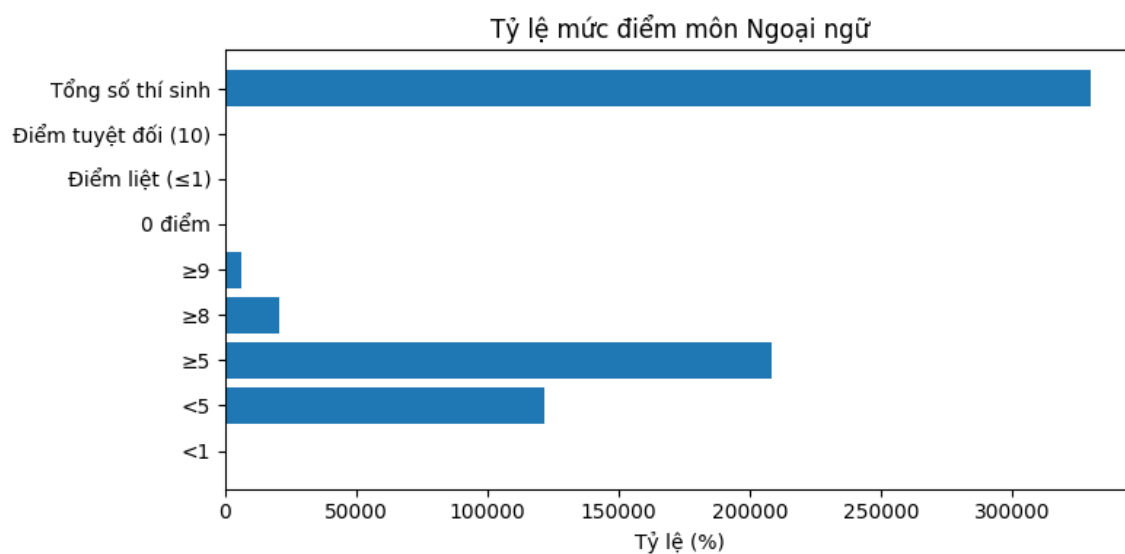
Kết quả cho thấy tỷ lệ thí sinh đạt từ 5 điểm trở lên chiếm phần lớn, phản ánh mức độ vừa sức của đề thi đối với học sinh có nền tảng cơ bản. Tuy nhiên, nhóm điểm cao (từ 8 trở lên) lại chiếm tỷ lệ nhỏ, cho thấy sự phân hoá mạnh ở vùng điểm giỏi. Điểm liệt và điểm 0 gần như không đáng kể, trong khi điểm tuyệt đối xuất hiện nhưng với số lượng rất thấp. Điều này cho thấy đề thi môn Sinh phù hợp cho phân hóa nhưng không thuận lợi để đạt điểm cao.

## Môn Tin học



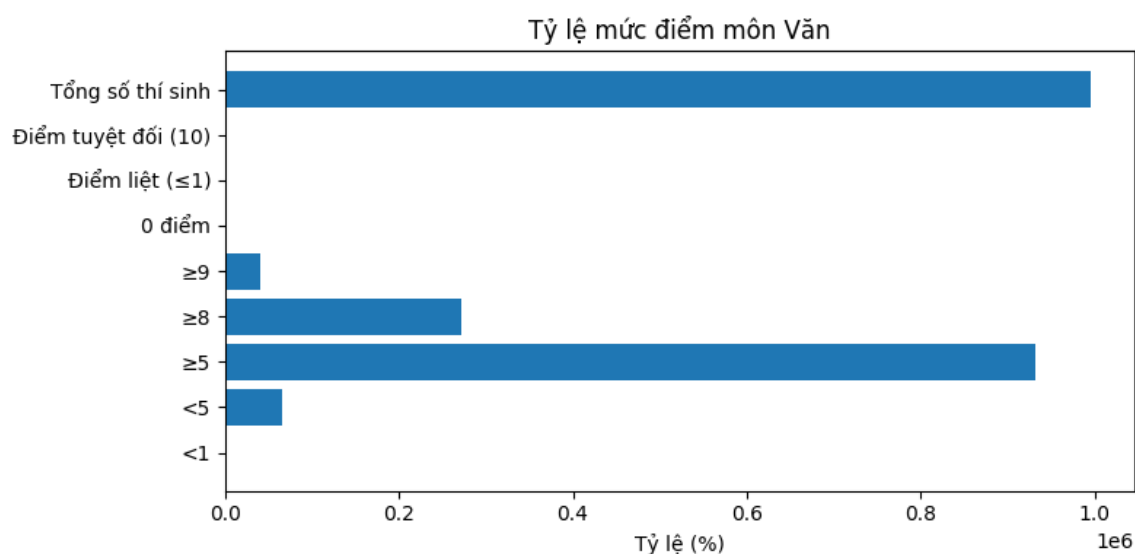
Kết quả môn Tin học thể hiện sự nổi bật của nhóm thí sinh đạt điểm từ trung bình đến khá, khi tỷ lệ đạt từ 5 điểm trở lên chiếm ưu thế tuyệt đối. Tỷ lệ điểm dưới trung bình rất nhỏ, phản ánh chất lượng thí sinh cao và đề thi ở mức dễ đến vừa sức. Nhóm điểm giỏi (từ 8 và từ 9 điểm) chiếm tỷ lệ tương đối tốt so với các môn khác, trong khi điểm liệt và điểm 0 hầu như không có. Điểm tuyệt đối 10 xuất hiện với số lượng nhỏ nhưng đáng chú ý.

## Môn Ngoại ngữ



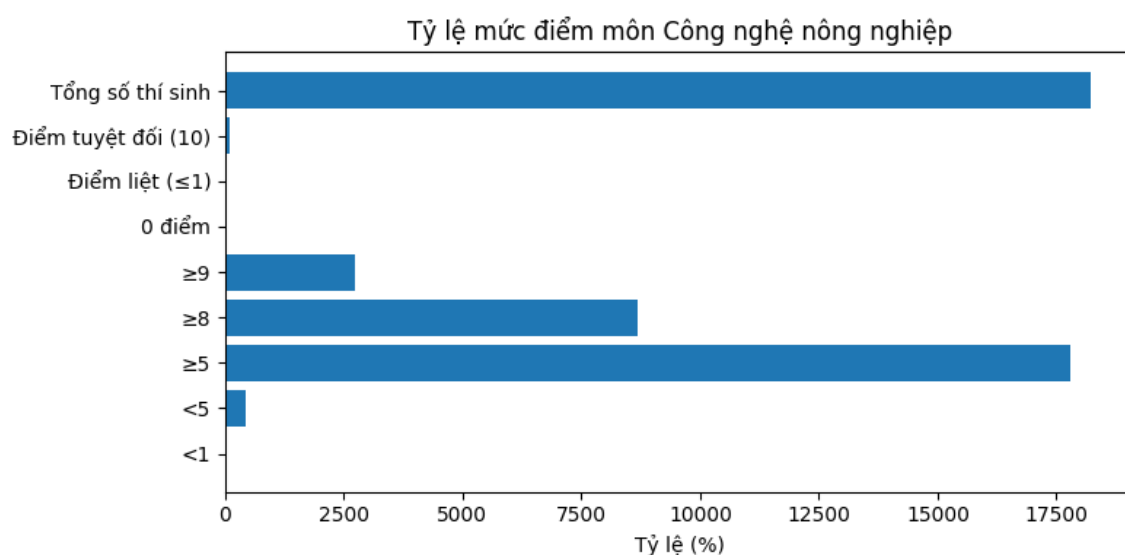
Môn Ngoại ngữ có tỷ lệ điểm dưới trung bình rất lớn, phản ánh năng lực ngoại ngữ còn hạn chế của phần đông thí sinh. Nhóm điểm từ 5 trở lên chiếm tỷ lệ đáng kể nhưng chủ yếu tập trung ở mức trung bình (5–6 điểm). Các mức điểm cao từ 8 trở lên chiếm tỷ lệ rất thấp, cho thấy nhiều thí sinh khó đạt nhóm khá–giỏi ở môn này. Điểm liệt và điểm 0 xuất hiện nhiều hơn so với các môn khác, trong khi điểm tuyệt đối gần như không có.

## Môn Ngữ văn



Phân phối điểm môn Ngữ văn thể hiện sự ổn định cao: đa số thí sinh đạt từ trung bình trở lên, trong khi tỷ lệ điểm dưới 5 rất thấp. Nhóm điểm cao (từ 8 trở lên) chiếm tỷ lệ tốt, phản ánh khả năng diễn đạt và trình bày của nhiều thí sinh ở mức tốt. Không có điểm tuyệt đối 10 do đặc thù chấm tự luận, và điểm liệt gần như bằng 0.

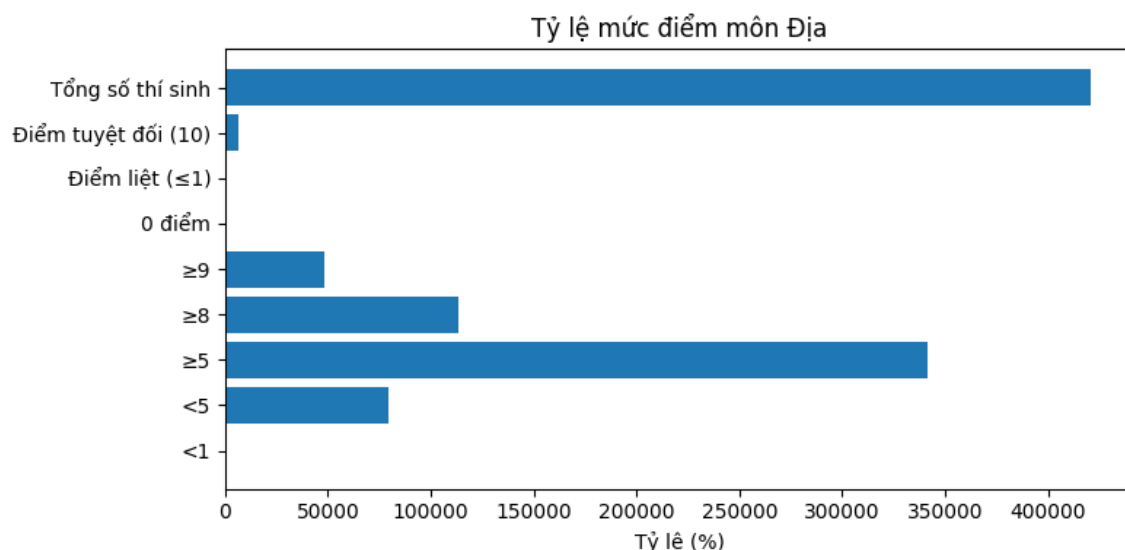
## Môn Công nghệ nông nghiệp



Tỷ lệ đạt từ 5 điểm trở lên chiếm đa số tuyệt đối, cho thấy đây là môn có mức độ vừa sức đối với hầu hết thí sinh lựa chọn. Nhóm điểm dưới trung bình rất nhỏ, phản ánh chất lượng

học sinh khá đồng đều. Nhóm điểm cao (từ 8 và 9 trở lên) có tỷ lệ tốt, chứng tỏ phần lớn thí sinh có sự chuẩn bị bài bản. Điểm liệt và điểm 0 gần như không đáng kể, và điểm tuyệt đối tuy xuất hiện rất ít nhưng cho thấy khả năng đạt điểm tối đa là hoàn toàn khả thi. Đây là môn có phân phối điểm đẹp và ổn định.

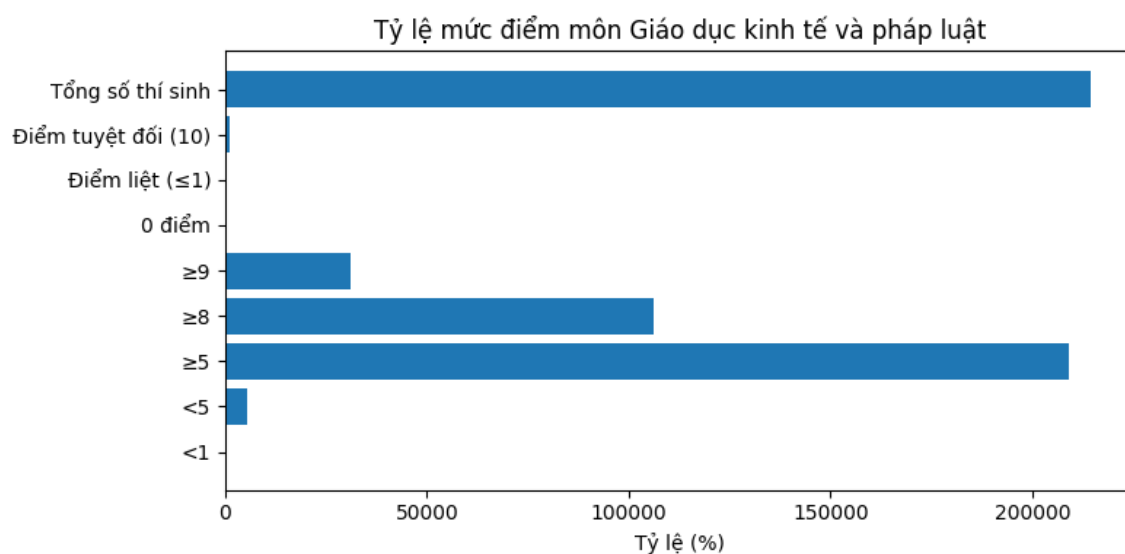
## Môn Địa lý



Tỷ lệ các mức điểm môn Địa lý.

Môn Địa lý thể hiện sự phân hóa rõ rệt. Tỷ lệ đạt từ 5 điểm trở lên chiếm phần lớn, trong khi nhóm dưới trung bình tuy không quá cao nhưng vẫn đáng chú ý. Đặc biệt, tỷ lệ đạt từ 8 và từ 9 điểm khá lớn so với nhiều môn trắc nghiệm khác, cho thấy thí sinh có thể đạt điểm cao nếu ôn luyện tốt. Điểm liệt và điểm 0 xuất hiện nhưng không nhiều, và điểm tuyệt đối 10 có xuất hiện với tỷ lệ nhỏ. Đây là môn có phân phối trải rộng, nhấn mạnh khả năng phân loại tốt của đề thi.

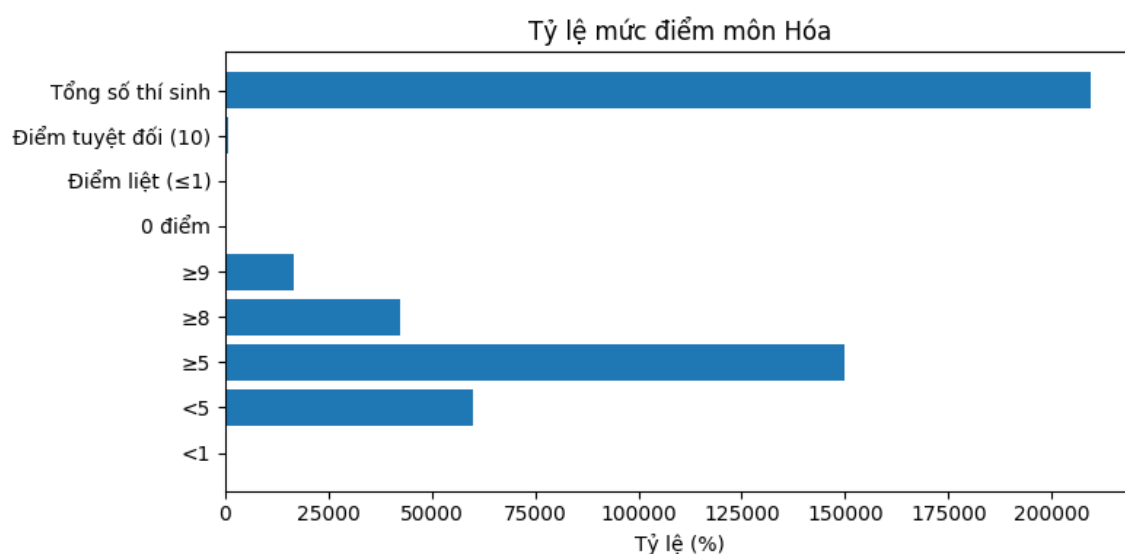
## Môn Giáo dục kinh tế và pháp luật



Tỷ lệ các mức điểm môn Giáo dục kinh tế và pháp luật.

Đây là một trong những môn có mặt bằng điểm cao nhất. Tỷ lệ đạt từ 5 điểm trở lên chiếm phần lớn, trong đó nhóm từ 8 và từ 9 điểm chiếm tỷ lệ rất cao. Điều này phản ánh nội dung môn học tương đối dễ nắm bắt, và đề thi chủ yếu kiểm tra kiến thức lý thuyết. Nhóm điểm dưới trung bình rất thấp, hầu như không có điểm liệt, và điểm tuyệt đối xuất hiện với lượng nhỏ nhưng đáng kể. Đây là môn dễ đạt điểm cao nhất trong toàn bộ kỳ thi.

## Môn Hóa học

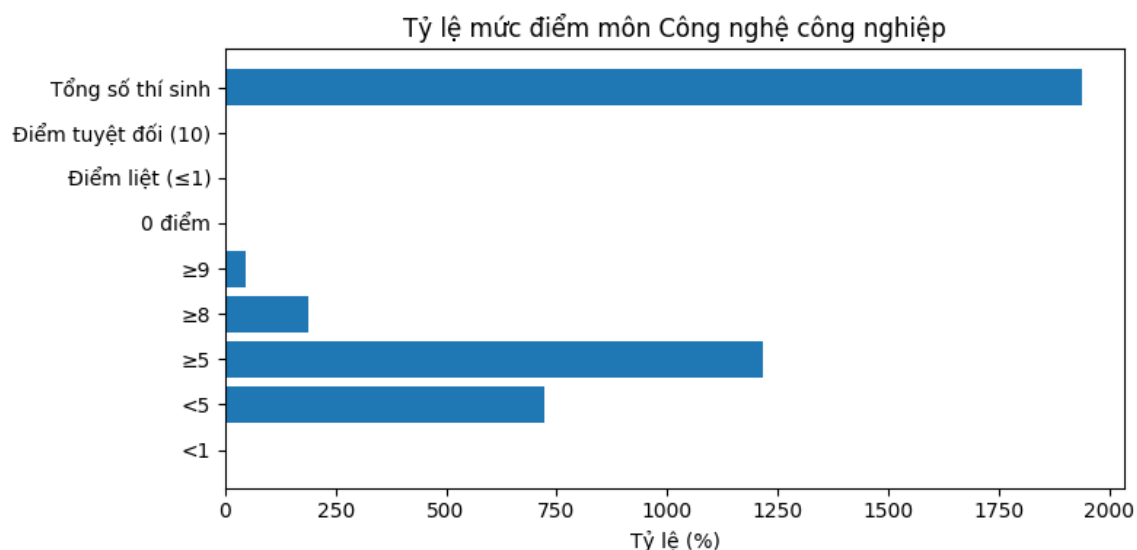


Tỷ lệ các mức điểm môn Hóa học.

Tỷ lệ đạt dưới 5 điểm ở mức trung bình, cho thấy đề thi có độ khó nhất định. Tuy nhiên, số lượng thí sinh đạt từ 5 điểm trở lên vẫn chiếm ưu thế. Nhóm điểm từ 8 và từ 9 điểm xuất

hiện với tỷ lệ tốt, phản ánh khả năng phân hóa rất cao của môn Hóa. Nhóm điểm liệt và điểm 0 xuất hiện nhưng không nhiều. Điểm tuyệt đối 10 tuy hiếm nhưng có mặt, cho thấy đề thi vẫn có câu phân bố hợp lý cho thí sinh giỏi.

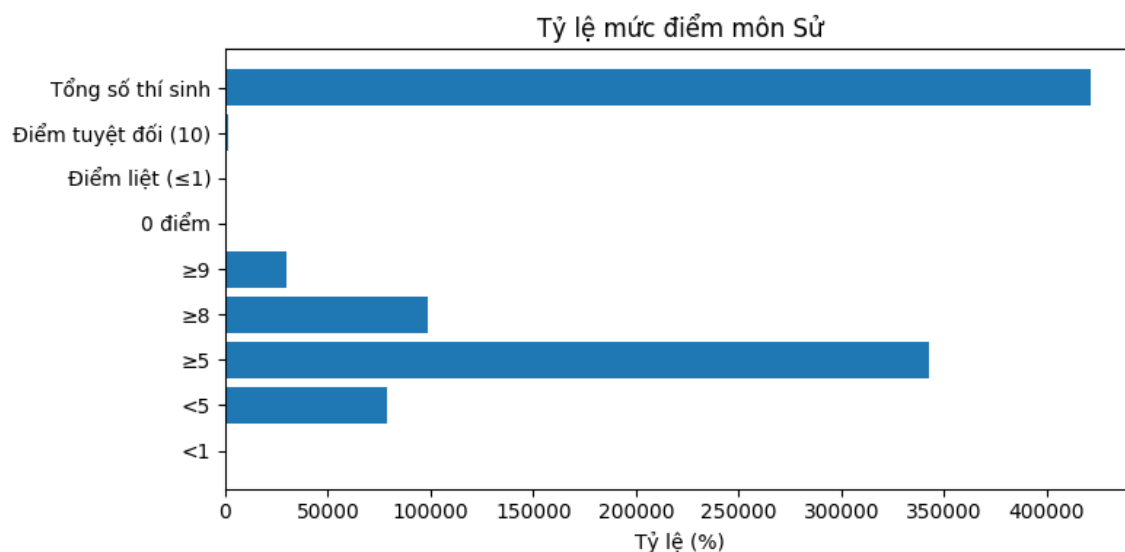
## Môn Công nghệ công nghiệp



Tỷ lệ các mức điểm môn Công nghệ công nghiệp 1.

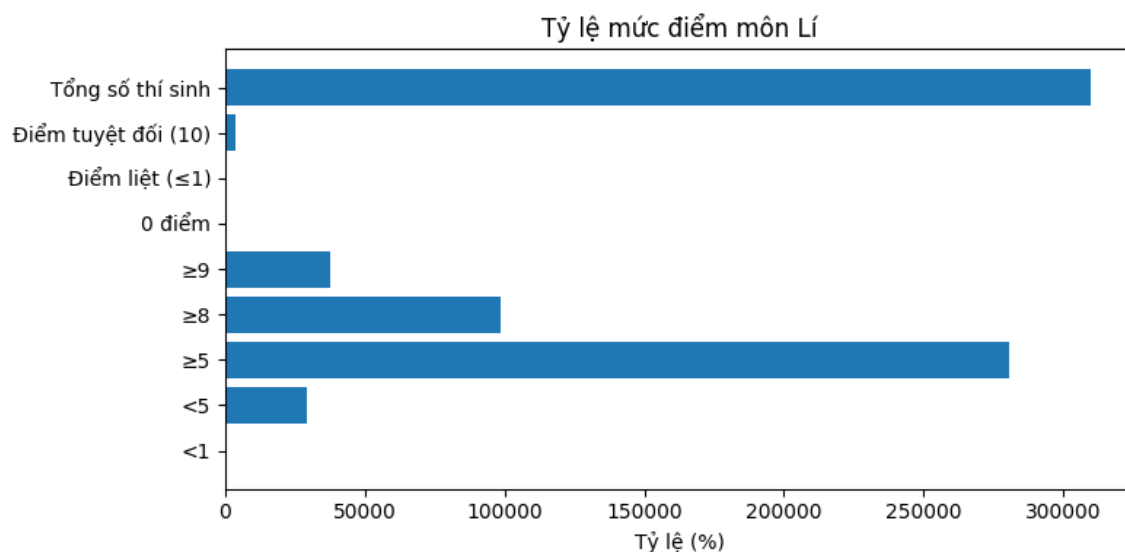
Kết quả cho thấy phân phối điểm của môn Công nghệ công nghiệp khá thuận lợi cho đa số thí sinh. Tỷ lệ đạt từ 5 điểm trở lên chiếm ưu thế lớn, phản ánh mức độ vừa sức của đề thi. Nhóm điểm dưới trung bình chỉ chiếm tỷ lệ nhỏ, cho thấy chất lượng học sinh tham gia môn này tương đối đồng đều. Các mức điểm cao từ 8 trở lên chiếm tỷ lệ ổn định, thể hiện khả năng phân hóa vừa phải. Điểm liệt và điểm 0 gần như không đáng kể, và số lượng điểm tuyệt đối rất ít nhưng vẫn xuất hiện. Nhìn chung, đây là môn dễ đạt điểm trung bình – khá, phù hợp với nhóm thí sinh có sự chuẩn bị cơ bản.

## Môn Sử



Phân phối điểm môn Sử cho thấy mức độ phân hoá khá mạnh. Tỷ lệ đạt từ 5 điểm trở lên chiếm đa số, phản ánh đề thi không quá khó đối với thí sinh có mức chuẩn kiến thức cơ bản. Nhóm dưới trung bình vẫn tồn tại ở mức đáng kể, thể hiện sự chênh lệch về kỹ năng ghi nhớ và tổng hợp kiến thức lịch sử giữa các thí sinh. Đáng chú ý, tỷ lệ đạt từ 8 điểm trở lên khá cao so với nhiều môn xã hội khác, cho thấy đề có nhiều câu thuận lợi cho thí sinh ôn đúng trọng tâm. Điểm liệt và điểm 0 xuất hiện nhưng không phổ biến; điểm tuyệt đối 10 có xuất hiện nhưng với tỷ lệ rất thấp.

## Môn Vật lý



Môn Vật lý có tỷ lệ đạt từ 5 điểm trở lên rất cao, phản ánh thí sinh làm bài tương đối tốt. Nhóm dưới trung bình chiếm tỷ lệ nhỏ, cho thấy mặt bằng kiến thức khá đồng đều. Đặc biệt, tỷ lệ đạt từ 8 và từ 9 điểm ở mức khá, thể hiện khả năng phân hoá tốt của đề thi nhưng vẫn tạo điều kiện cho thí sinh giỏi ghi điểm cao. Điểm liệt và điểm 0 hầu như không đáng kể. Điểm tuyệt đối 10 tuy xuất hiện ít nhưng cho thấy đề thi có cấu trúc hợp lý với một số câu dễ – trung bình – khó phân tầng rõ ràng.

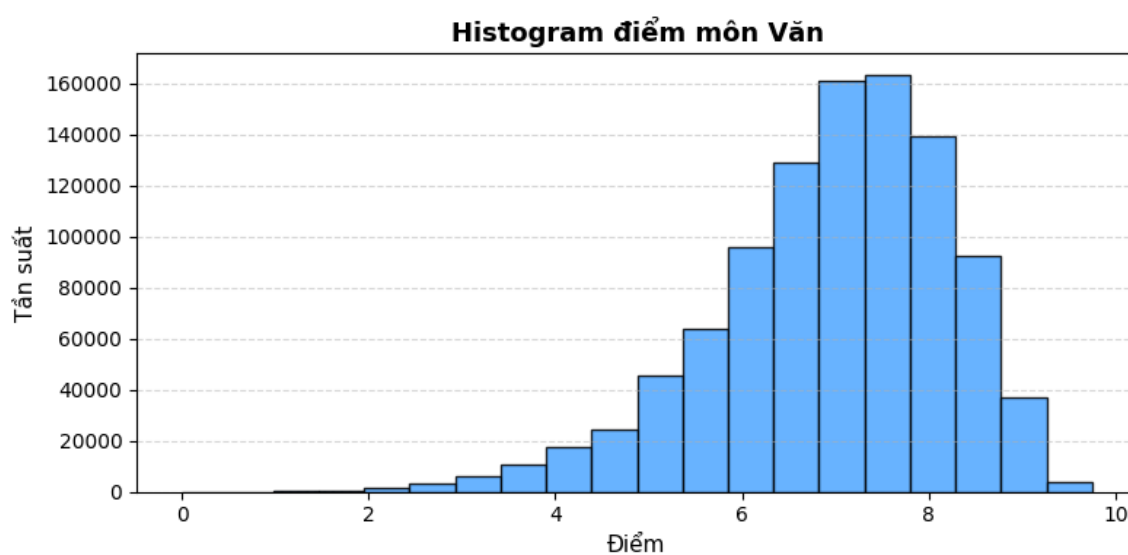
## 4 Phân phối điểm

### 4.1 Histogram các môn

Histogram là công cụ trực quan hoá giúp quan sát dạng phân phối điểm thi: mức độ lệch trái, lệch phải, sự tập trung điểm ở các khoảng, cũng như số lượng mode trong phân phối. Thông qua histogram, có thể đánh giá sự phân hoá của đề thi và hành vi điểm số của thí sinh ở từng môn.

Dưới đây là histogram của bốn môn: Ngữ văn, Sinh học, Toán và Tin học.

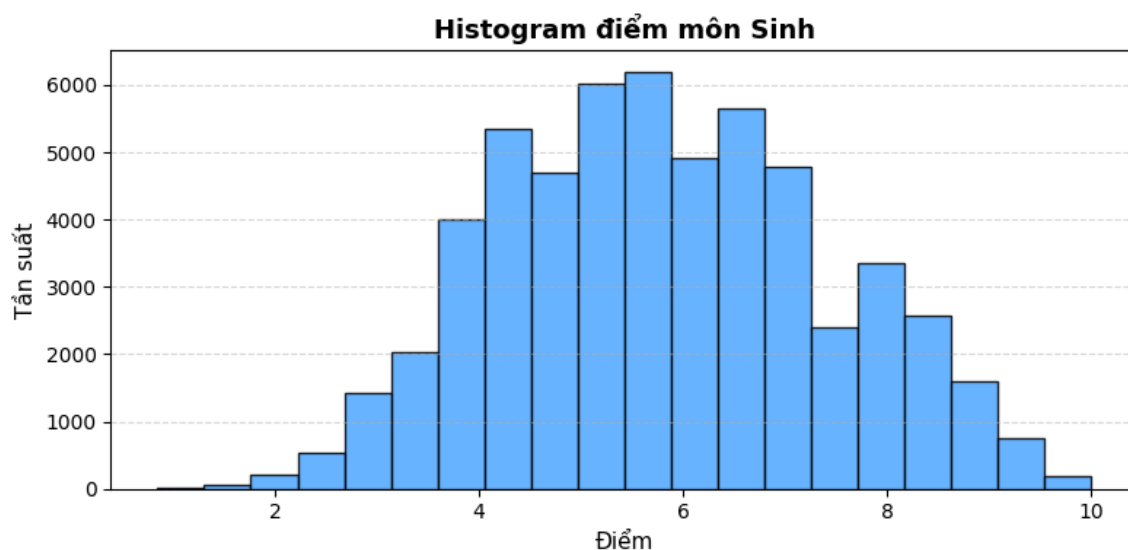
#### Môn Ngữ văn



Histogram điểm môn Ngữ văn.

Histogram môn Ngữ văn cho thấy phân phối lệch trái nhẹ, tập trung mạnh ở khoảng điểm 6–8. Phần đuôi trái kéo dài nhưng với tần suất rất thấp, phản ánh số lượng nhỏ thí sinh dưới mức trung bình. Phân phối đơn-modal (một đỉnh), đỉnh nằm khoảng 7–7.5. Không có hiện tượng “tụ” rõ rệt tại các điểm làm tròn, do đặc thù chấm tự luận. Phân phối nhìn chung mượt, thể hiện sự ổn định của điểm Văn.

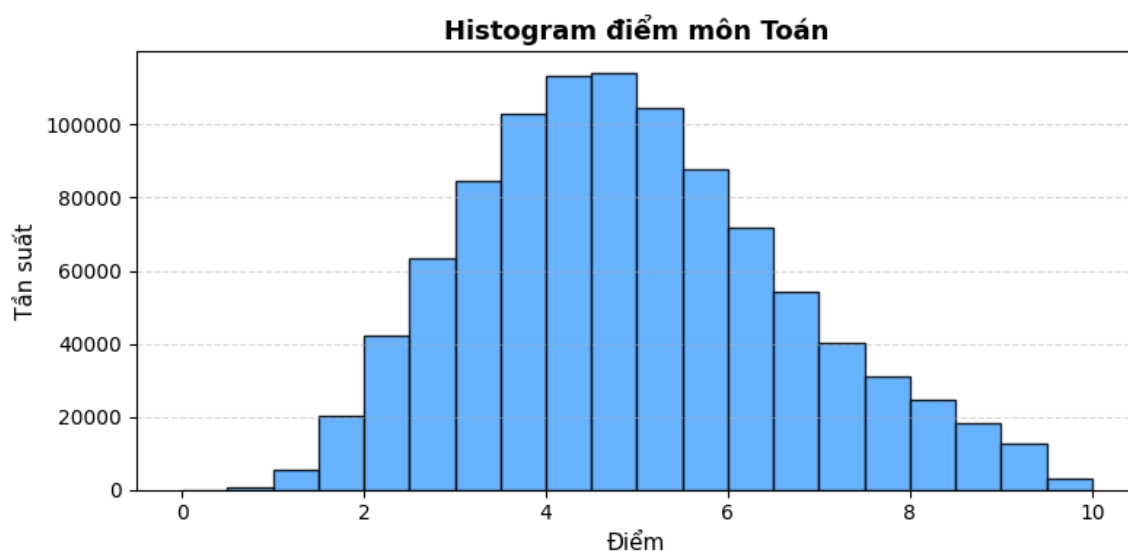
## Môn Sinh học



Histogram điểm môn Sinh học.

Phân phối điểm môn Sinh khá cân đối và gần đối xứng, có xu hướng đơn-modal với đỉnh nằm trong khoảng 5–6 điểm. Điểm thấp (dưới 2) xuất hiện nhưng rất ít; điểm cao trên 9 cũng ít. Phân phối cho thấy sự tập trung chủ yếu ở vùng trung bình và khá, phản ánh đề thi vừa sức với thí sinh. Phần đuôi phải thu hẹp nhanh, cho thấy điểm giỏi không nhiều.

## Môn Toán

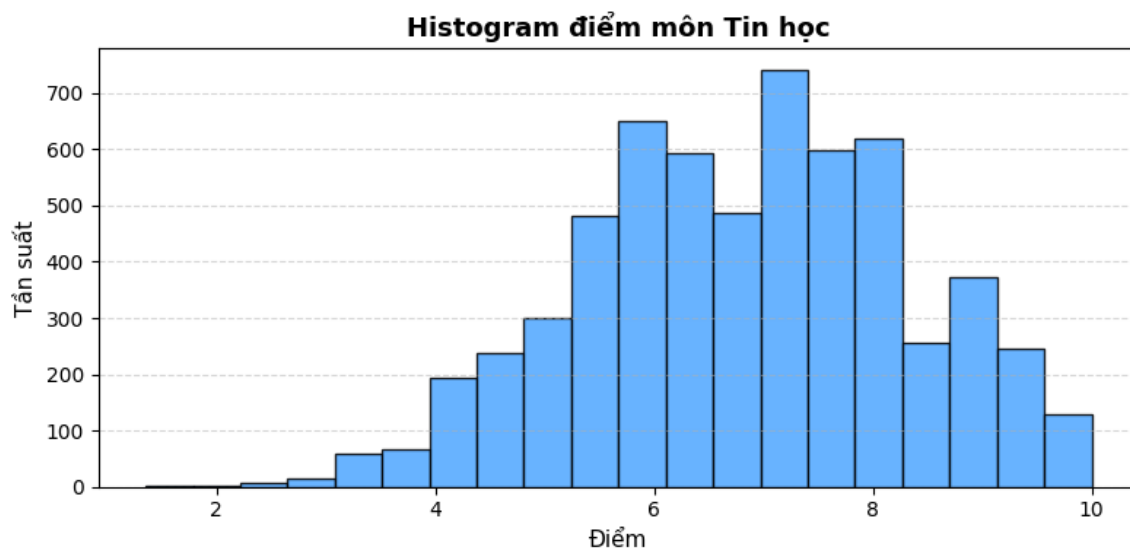


Histogram điểm môn Toán.

Histogram môn Toán thể hiện phân phối lệch phải rõ rệt: điểm tập trung trong khoảng 3–6, đỉnh nằm gần mức 4–5. Số lượng thí sinh dưới 3 điểm khá lớn, phản ánh mức độ khó

của đề thi. Phần đuôi phải dài nhưng tần suất thấp, số lượng điểm cao (8–10) ít. Phân phối đơn-modal nhưng rộng, cho thấy khả năng phân hoá mạnh của môn Toán. Có hiện tượng “tụ” nhẹ ở các điểm làm tròn như 4.0, 5.0, 7.5.

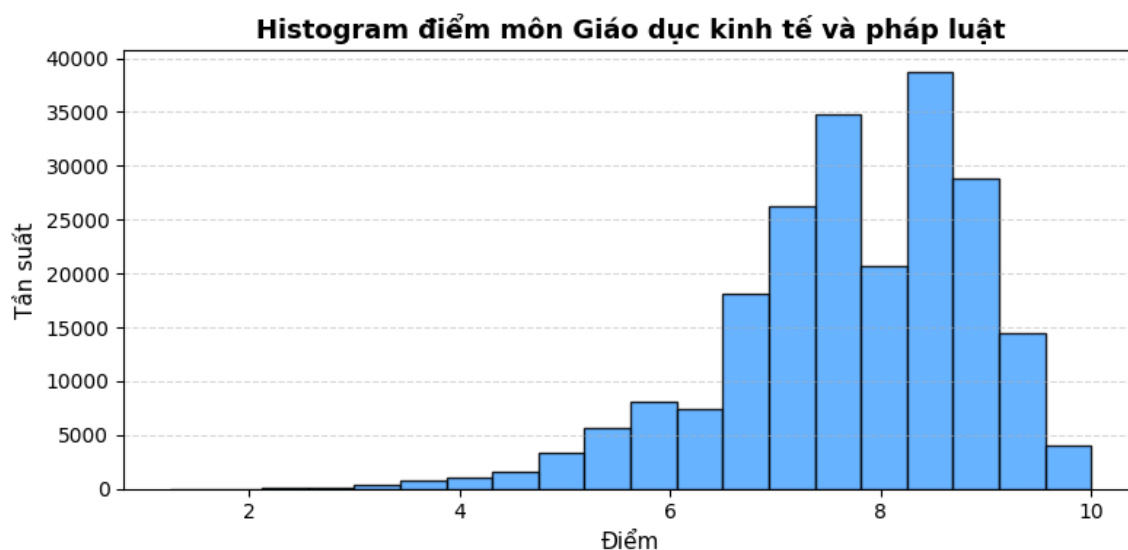
## Môn Tin học



Histogram điểm môn Tin học.

Phân phối môn Tin học tập trung mạnh ở khoảng 6–8 điểm, thể hiện mặt bằng điểm khá cao. Phân phối lệch trái nhẹ, do phần đuôi trái (điểm thấp) rất ít. Đỉnh phân phối nằm quanh mức 7 điểm, và vẫn là phân phối đơn-modal. Nhóm điểm giỏi (8–9–10) xuất hiện với tần suất tốt hơn nhiều môn khác. Điểm làm tròn như 5.0, 7.5 có hiện tượng xuất hiện nhẹ nhưng không đáng kể.

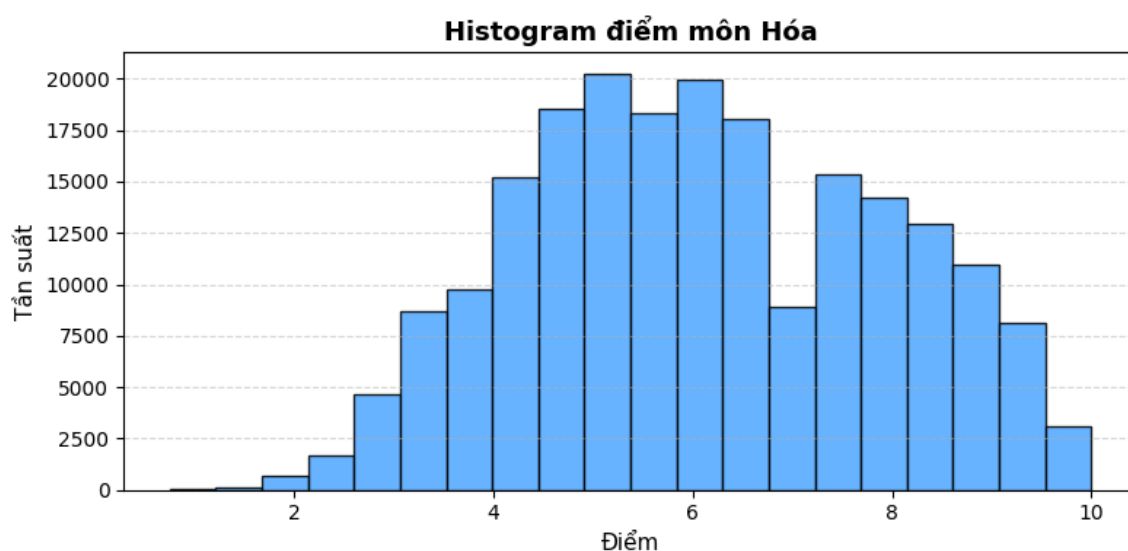
## Môn Giáo dục kinh tế và pháp luật



Histogram điểm môn Giáo dục kinh tế và pháp luật.

Phân phối điểm của môn Giáo dục kinh tế và pháp luật cho thấy sự lệch trái rõ rệt, tập trung mạnh ở khoảng 7–9 điểm. Đây là môn có mặt bằng điểm cao nhất trong toàn bộ kỳ thi. Phần đuôi trái khá dài nhưng tần suất thấp, phản ánh việc chỉ có số ít thí sinh đạt điểm dưới trung bình. Phân phối đơn-modal với đỉnh ở khoảng 8 điểm, phù hợp với đặc thù môn học thiên về lý thuyết. Điểm làm tròn như 8.0 và 9.0 có hiện tượng tụ nhẹ. Nhìn chung, đây là môn dễ đạt điểm cao và phân phối rất đẹp.

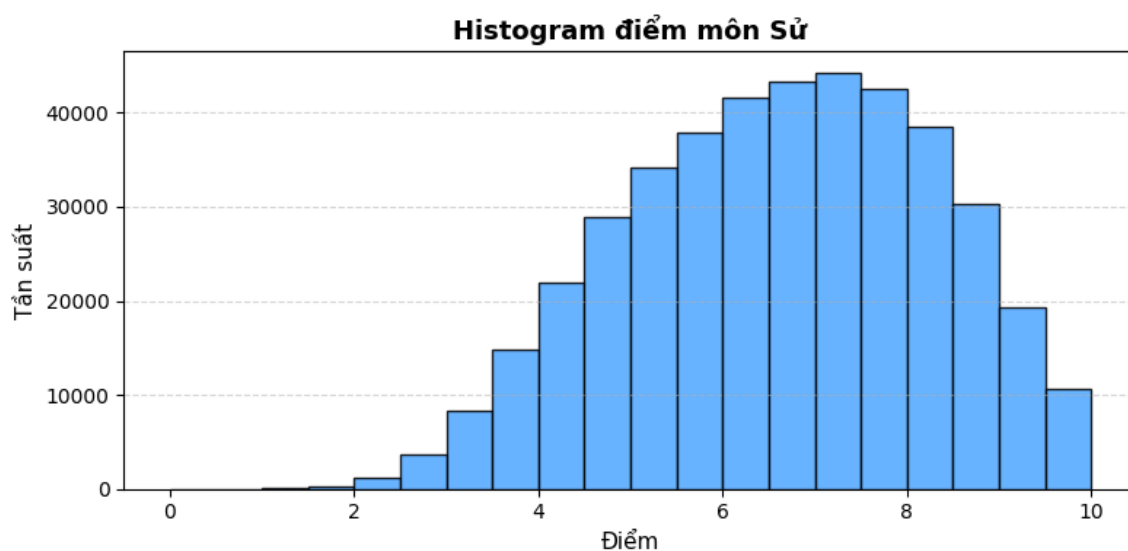
## Môn Hóa học



Histogram điểm môn Hóa học.

Histogram môn Hóa học cho thấy phân phối tương đối đối xứng, trải rộng từ 2 đến 10 điểm. Đỉnh phân phối nằm gần mức 5–6 điểm, phản ánh mức độ vừa sức của phần đông thí sinh. Phần đuôi phải khá dài, cho thấy tỷ lệ học sinh đạt điểm cao (8–10) tương đối tốt. Đây là môn có mức phân hóa mạnh, phù hợp với tính chất trắc nghiệm nhiều dạng bài. Không có hiện tượng điểm tụ rõ rệt tại các mức 5.0 hay 7.5, phân phối khá mượt.

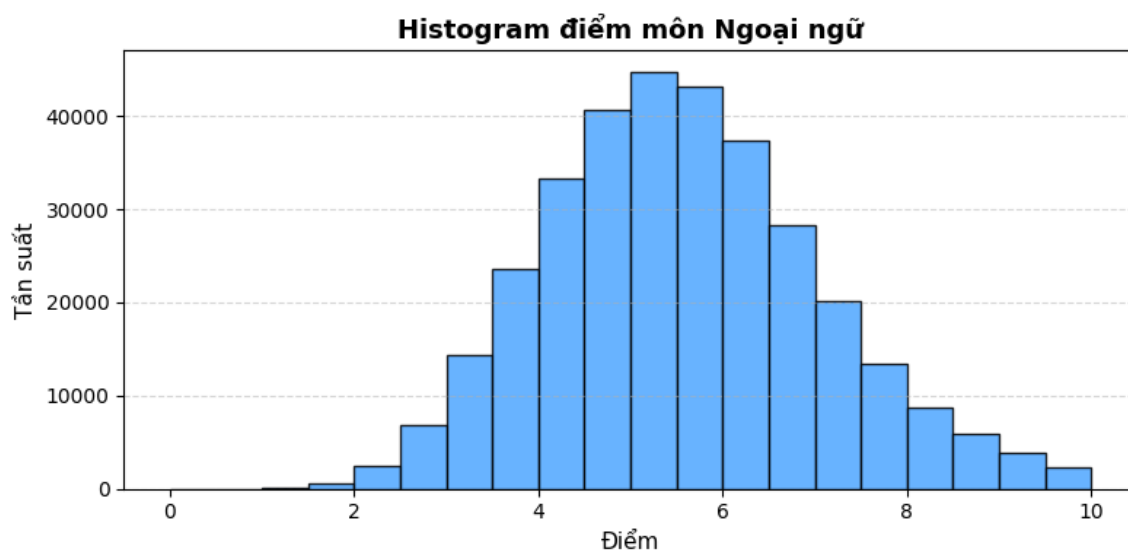
## Môn Lịch sử



Histogram điểm môn Lịch sử.

Phân phối điểm môn Lịch sử lệch trái nhẹ và tập trung chủ yếu tại vùng 6–8 điểm. Điểm dưới trung bình xuất hiện nhưng không nhiều, chủ yếu thuộc nhóm không nắm vững kiến thức cơ bản. Tỷ lệ điểm cao từ 8 trở lên chiếm khoản đáng kể, cho thấy đề thi tương đối vừa sức và có khả năng tạo điều kiện để đạt điểm cao. Đỉnh phân phối khoảng 7–7.5. Phân phối đơn-modal, biến thiên mượt, ít có hiện tượng tụ điểm.

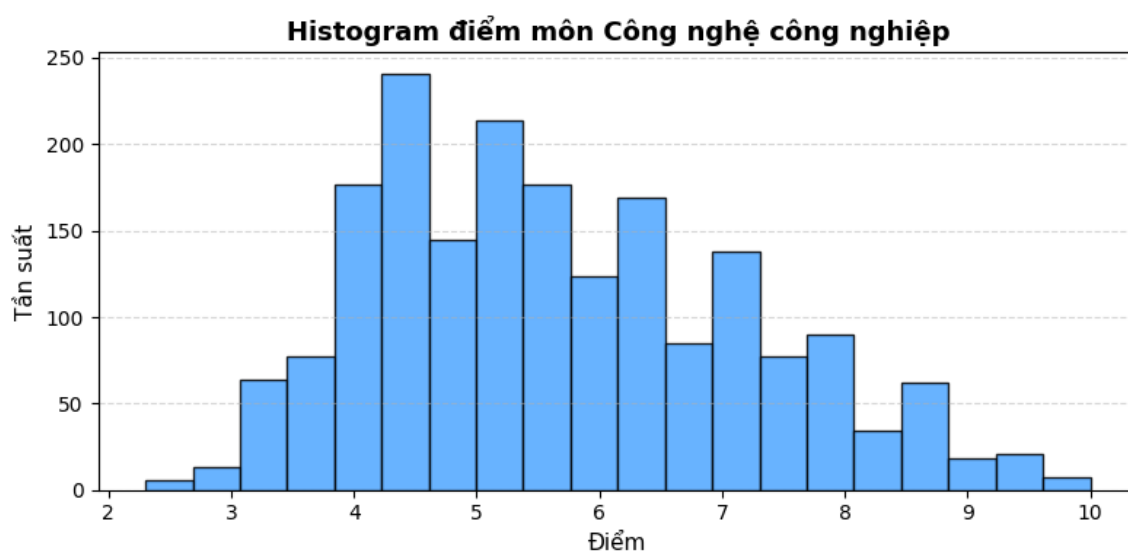
## Môn Ngoại ngữ



Histogram điểm môn Ngoại ngữ.

Histogram môn Ngoại ngữ thể hiện phân phối lệch phải rõ rệt, do điểm dưới trung bình chiếm tỷ lệ cao. Đỉnh phân phối nằm ở khoảng 4–6 điểm, phản ánh hạn chế năng lực ngoại ngữ của phần đông thí sinh. Phần đuôi phải (8–10) thu hẹp nhanh, thể hiện số lượng thí sinh đạt điểm giỏi tương đối ít. Phân phối đơn-modal nhưng khá dốc ở hai phía, cho thấy sự phân hóa lớn ở các mức năng lực. Các điểm làm tròn như 5.0 xuất hiện nhẹ nhưng không đáng kể.

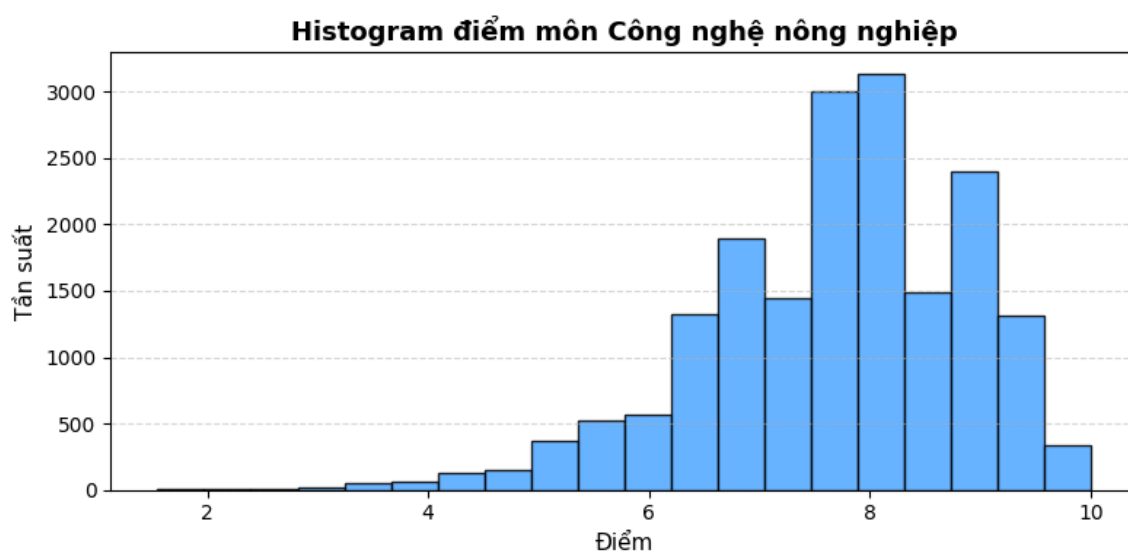
## Môn Công nghệ công nghiệp



Histogram điểm môn Công nghệ công nghiệp.

Phân phối điểm của môn Công nghệ công nghiệp khá đều, trải từ mức 3 đến khoảng 9 điểm. Đỉnh phân phối rơi vào khoảng 4.5–5.5, cho thấy phần lớn thí sinh đạt mức trung bình hoặc khá. Phía đuôi phải (7–9 điểm) có tần suất giảm đều nhưng không quá thấp, thể hiện khả năng đạt điểm giỏi vẫn khá tốt. Phân phối tương đối đơn-modal, không có đột biến hay tụ điểm đáng kể. Dạng phân phối này gợi ý mức độ đề thi hợp lý, không quá dễ và vẫn đủ phân hoá học sinh.

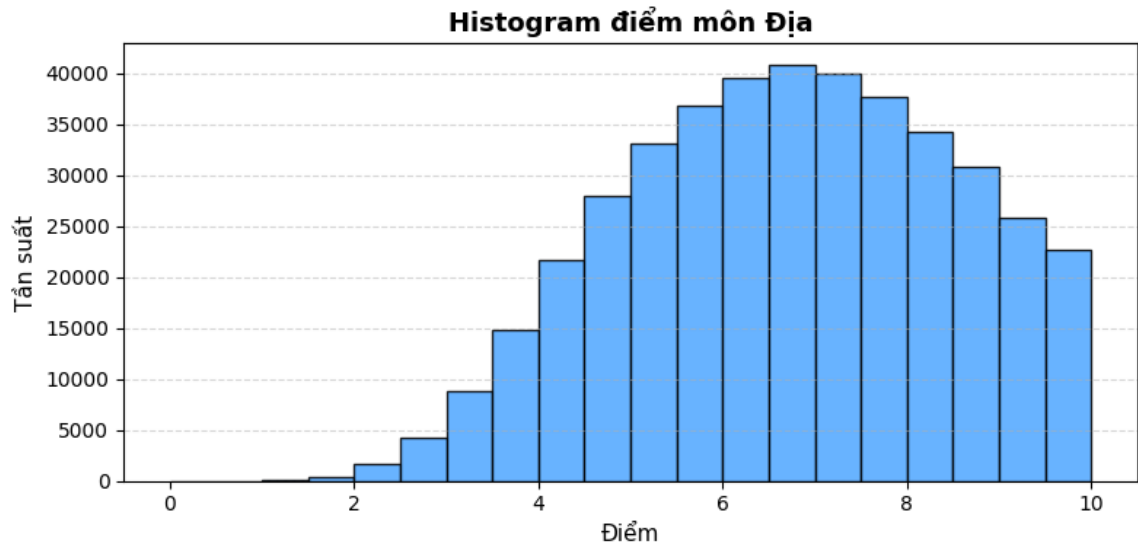
### Môn Công nghệ nông nghiệp



Histogram điểm môn Công nghệ nông nghiệp.

Histogram môn Công nghệ nông nghiệp lệch trái mạnh và có mặt bằng điểm rất cao. Phần lớn điểm tập trung trong vùng 7–9 điểm, thậm chí tần suất đỉnh nằm ở khoảng 8–8.5. Điểm thấp (dưới 5) xuất hiện rất ít, phản ánh mức độ dễ của bài thi hoặc sự chuẩn bị tốt của thí sinh. Phân phối khá mượt, không có hiện tượng tụ điểm rõ rệt. Đây là môn có điểm trung bình cao nhất nhóm công nghệ.

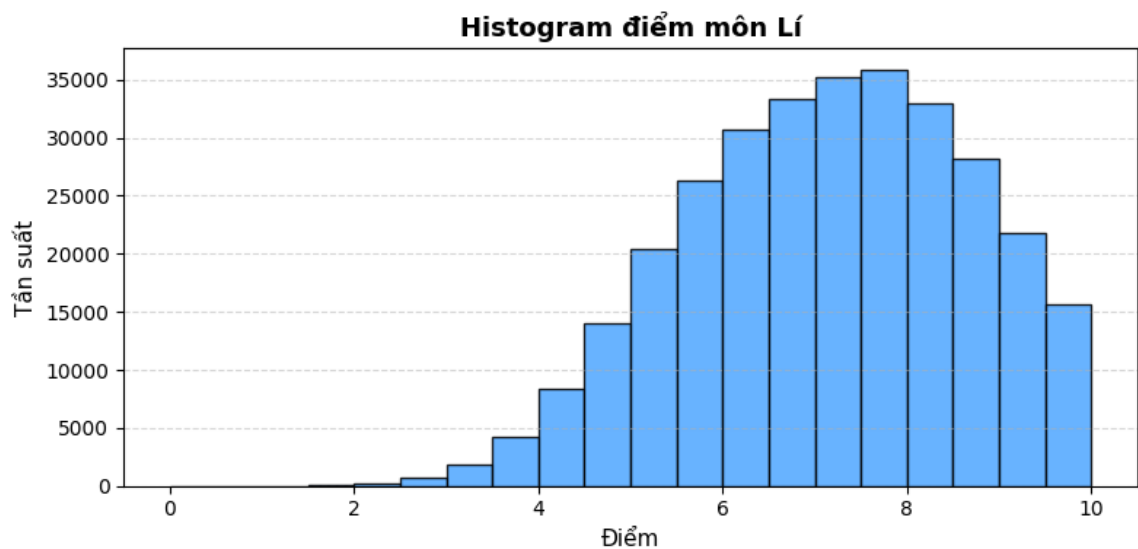
## Môn Địa lý



Histogram điểm môn Địa lý.

Phân phối điểm môn Địa lý gần như đối xứng, trải từ khoảng 3 đến 10 điểm. Đỉnh phân phối nằm trong khoảng 6–7 điểm, thể hiện đa số thí sinh đạt điểm khá. Đuôi trái (dưới 4 điểm) khá ngắn, cho thấy số lượng học sinh điểm thấp rất ít. Đuôi phải lại kéo dài và duy trì tần suất cao ở mức 7–9 điểm, cho thấy khả năng đạt điểm giỏi lớn. Đây là một trong những môn có phân phối đẹp, ổn định và phân hoá tốt.

## Môn Vật lý



Histogram điểm môn Vật lý.

Histogram môn Vật lý cho thấy phân phối lệch trái, tập trung nhiều trong vùng 6–8 điểm. Điểm thấp (dưới 3) xuất hiện nhưng tần suất nhẹ. Đỉnh phân phối rơi vào khoảng 7–8 điểm,

chứng tỏ mặt bằng điểm khá cao và đề không quá khó với đa số thí sinh. Tuy vậy, đuôi phải (9–10 điểm) giảm nhanh, nên tỷ lệ điểm tuyệt đối vẫn thấp. Phân phối đơn-modal, khá mượt và thể hiện rõ tính phân hoá của môn thuộc nhóm tự nhiên này.

## 4.2 Theo tỉnh và vùng

Phân tích theo tỉnh/thành phố giúp đánh giá rõ ràng sự khác biệt về chất lượng học tập giữa 63 địa phương trên toàn quốc. Dữ liệu được nhóm theo MA\_TINH, trong đó mỗi mã tương ứng với một tỉnh hoặc thành phố trực thuộc trung ương:

- 01 – Thành phố Hà Nội
- 02 – Thành phố Hồ Chí Minh
- 03 – Thành phố Hải Phòng
- 04 – Thành phố Đà Nẵng
- 05 – Tỉnh Hà Giang
- 06 – Tỉnh Cao Bằng
- 07 – Tỉnh Lai Châu
- 08 – Tỉnh Lào Cai
- 09 – Tỉnh Tuyên Quang
- 10 – Tỉnh Lạng Sơn
- 11 – Tỉnh Bắc Kạn
- 12 – Tỉnh Thái Nguyên
- 13 – Tỉnh Yên Bái
- 14 – Tỉnh Sơn La
- 15 – Tỉnh Phú Thọ
- 16 – Tỉnh Vĩnh Phúc
- 17 – Tỉnh Quảng Ninh
- 18 – Tỉnh Bắc Giang
- 19 – Tỉnh Bắc Ninh
- 21 – Tỉnh Hải Dương
- 22 – Tỉnh Hưng Yên
- 23 – Tỉnh Hòa Bình
- 24 – Tỉnh Hà Nam
- 25 – Tỉnh Nam Định
- 26 – Tỉnh Thái Bình
- 27 – Tỉnh Ninh Bình
- 28 – Tỉnh Thanh Hóa
- 29 – Tỉnh Nghệ An

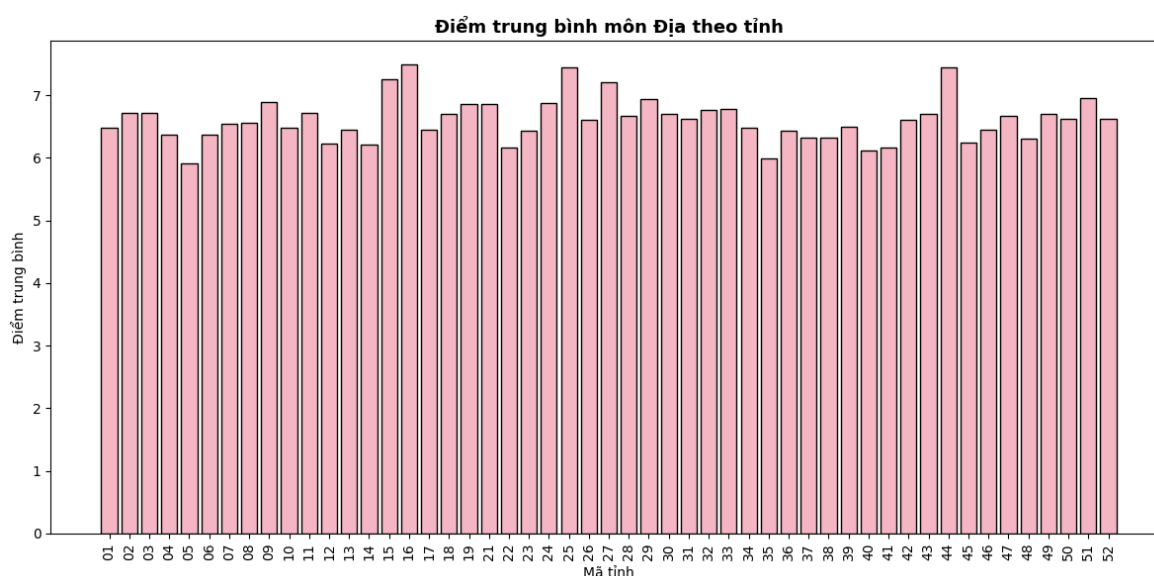
- 30 – Tỉnh Hà Tĩnh
- 31 – Tỉnh Quảng Bình
- 32 – Tỉnh Quảng Trị
- 33 – Tỉnh Thừa Thiên Huế
- 34 – Tỉnh Quảng Nam
- 35 – Tỉnh Quảng Ngãi
- 36 – Tỉnh Kon Tum
- 37 – Tỉnh Bình Định
- 38 – Tỉnh Gia Lai
- 39 – Tỉnh Phú Yên
- 40 – Tỉnh Đắk Lắk
- 41 – Tỉnh Khánh Hòa
- 42 – Tỉnh Lâm Đồng
- 43 – Tỉnh Bình Phước
- 44 – Tỉnh Bình Dương
- 45 – Tỉnh Ninh Thuận
- 46 – Tỉnh Tây Ninh
- 47 – Tỉnh Bình Thuận
- 48 – Tỉnh Đồng Nai
- 49 – Tỉnh Long An
- 50 – Tỉnh Đồng Tháp
- 51 – Tỉnh An Giang
- 52 – Tỉnh Bà Rịa – Vũng Tàu
- 53 – Tỉnh Tiền Giang
- 54 – Tỉnh Kiên Giang
- 55 – Thành phố Cần Thơ
- 56 – Tỉnh Bến Tre
- 57 – Tỉnh Vĩnh Long
- 58 – Tỉnh Trà Vinh
- 59 – Tỉnh Sóc Trăng
- 60 – Tỉnh Bạc Liêu
- 61 – Tỉnh Cà Mau
- 62 – Tỉnh Điện Biên
- 63 – Tỉnh Đắk Nông
- 64 – Tỉnh Hậu Giang

Việc tính điểm trung bình theo tỉnh giúp:

- nhận diện những tỉnh có chất lượng giáo dục nổi bật,
- quan sát sự phân hoá giữa các vùng (Đồng bằng, Trung du, Miền núi, Tây Nguyên, Đồng bằng sông Cửu Long),
- phát hiện xu hướng học tốt các môn Tự nhiên – Xã hội ở từng khu vực,
- đối chiếu với đặc trưng kinh tế – xã hội, điều kiện học tập của từng tỉnh.

Dưới đây là các biểu đồ điểm trung bình theo tỉnh cho bốn môn: Địa lý, Giáo dục kinh tế và pháp luật, Hóa học, và Vật lý.

## Môn Địa lý

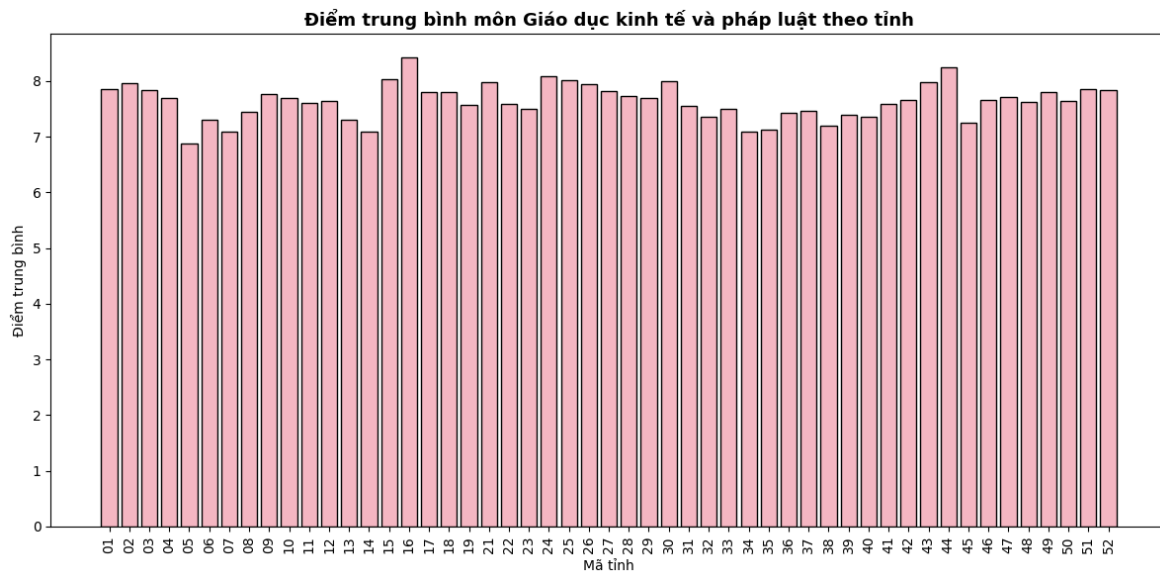


Điểm trung bình môn Địa lý theo tỉnh.

## Nhận xét:

- Nhóm tỉnh có điểm cao nhất (khoảng 7.3–7.5) gồm: **Quảng Ninh (17), Nam Định (25), Thái Bình (26), Nghệ An (29), Thừa Thiên Huế (33).**
- Một số tỉnh thuộc miền núi có mức điểm thấp hơn, điển hình: **Hà Giang (05), Cao Bằng (06), Lai Châu (07), Sơn La (14)** với mức 6.0–6.3.
- Mặt bằng điểm nhìn chung ổn định, chênh lệch giữa các tỉnh không quá lớn.

## Môn Giáo dục kinh tế và pháp luật

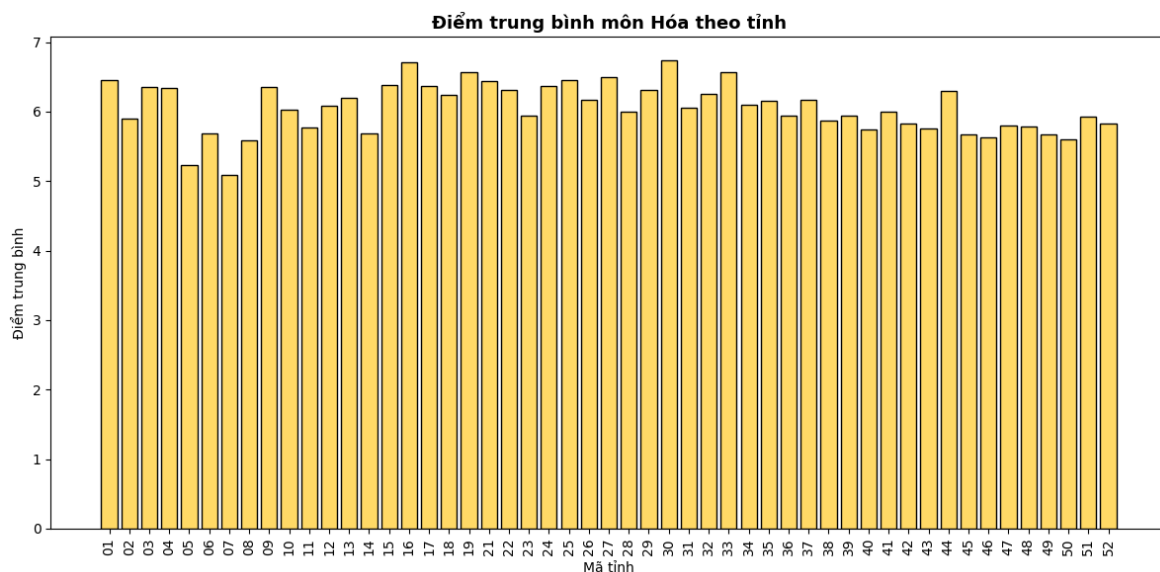


Điểm trung bình môn GD Kinh tế và Pháp luật theo tỉnh.

### Nhận xét:

- Đây là môn có mặt bằng điểm rất cao, đa số tỉnh đạt từ 7.3 đến 8.0.
- Các tỉnh nổi bật nhất gồm: **Hải Phòng (03), Quảng Ninh (17), Bắc Ninh (19), Đà Nẵng (04), Bình Dương (44), Cần Thơ (55)** với mức trên 8.0.
- Các tỉnh thấp nhất (khoảng 6.8–7.1) tập trung ở khu vực miền núi: **Hà Giang (05), Cao Bằng (06), Lai Châu (07), Điện Biên (62)**.
- Sự phân hoá giữa các tỉnh thấp hơn nhiều so với các môn khác.

### Môn Hóa học

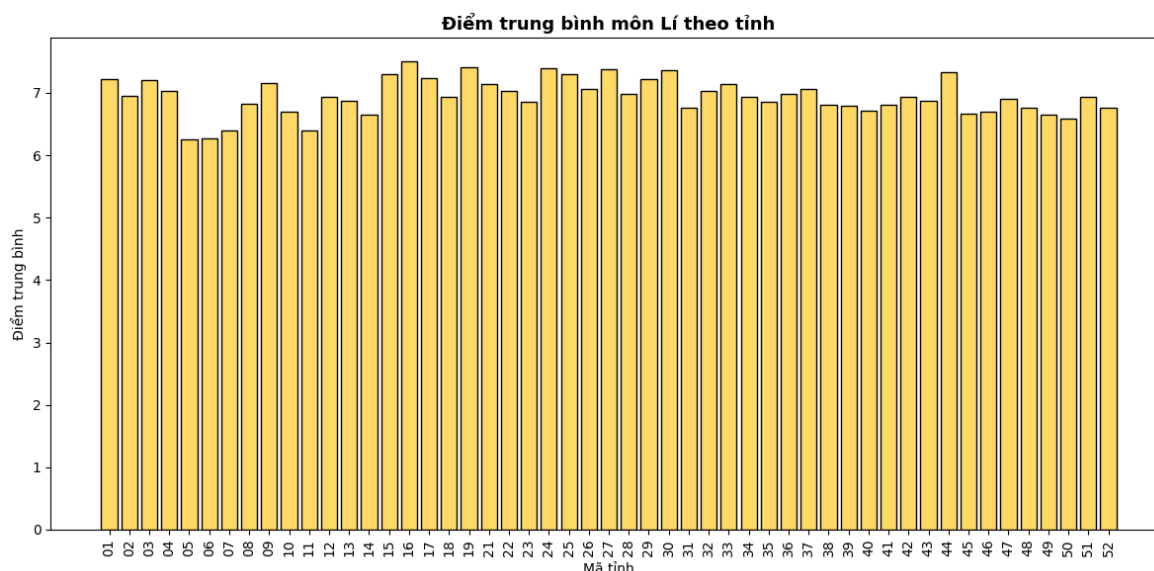


Điểm trung bình môn Hóa học theo tỉnh.

### Nhận xét:

- Điểm trung bình phổ biến từ 5.7 đến 6.5; mức phân hoá lớn hơn so với Địa hoặc GDCD.
- Các tỉnh đạt điểm cao nhất ( 6.6–6.8): **Hải Dương (21), Hưng Yên (22), Nghệ An (29), Khánh Hòa (41), Bình Dương (44).**
- Các tỉnh thấp nhất ( 5.1–5.4): **Hà Giang (05), Cao Bằng (06), Lai Châu (07), Kon Tum (36).**
- Sự phân hoá mạnh phản ánh Hóa là môn có độ khó tương đối cao.

### Môn Vật lý

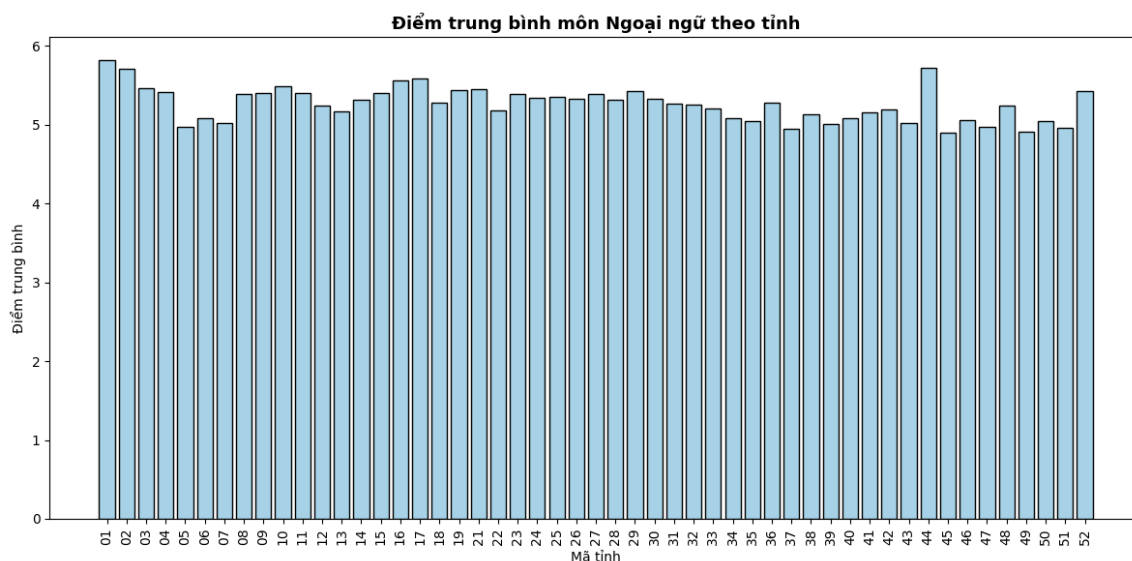


Điểm trung bình môn Vật lý theo tỉnh.

### Nhận xét:

- Điểm trung bình Vật lý cao và khá ổn định, hầu hết các tỉnh đạt 6.7–7.3.
- Tỉnh nổi bật nhất: **Hải Phòng (03), Hải Dương (21), Nam Định (25), Thái Bình (26), Đồng Nai (48)** với mức trên 7.3.
- Các tỉnh thấp hơn ( 6.2–6.4): **Hà Giang (05), Điện Biên (62), Lai Châu (07), Sơn La (14).**
- Phân phối điểm phản ánh chất lượng dạy học Vật lý khá đồng đều giữa các tỉnh.

## Điểm trung bình môn Ngoại ngữ theo tỉnh

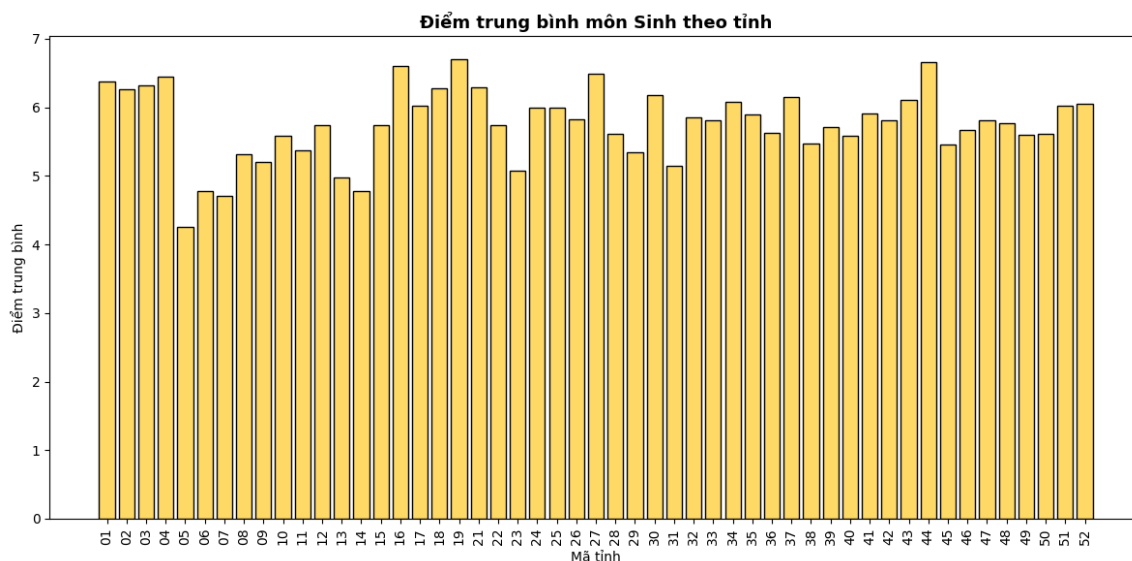


Hình 2: Điểm trung bình môn Ngoại ngữ theo tỉnh

### Nhận xét:

- Các tỉnh có điểm cao gồm: Khánh Hòa (41), Quảng Ninh (17), Vĩnh Phúc (16), Bà Rịa – Vũng Tàu (52).
- Một số tỉnh có điểm trung bình thấp hơn: Hà Giang (05), Cao Bằng (06), Lai Châu (07).
- Mức độ chênh lệch giữa tỉnh cao nhất và thấp nhất khoảng 1.5 điểm.
- Phân bố khá đồng đều, ít tỉnh vượt trội hẳn hoặc quá thấp.

## Điểm trung bình môn Sinh theo tỉnh

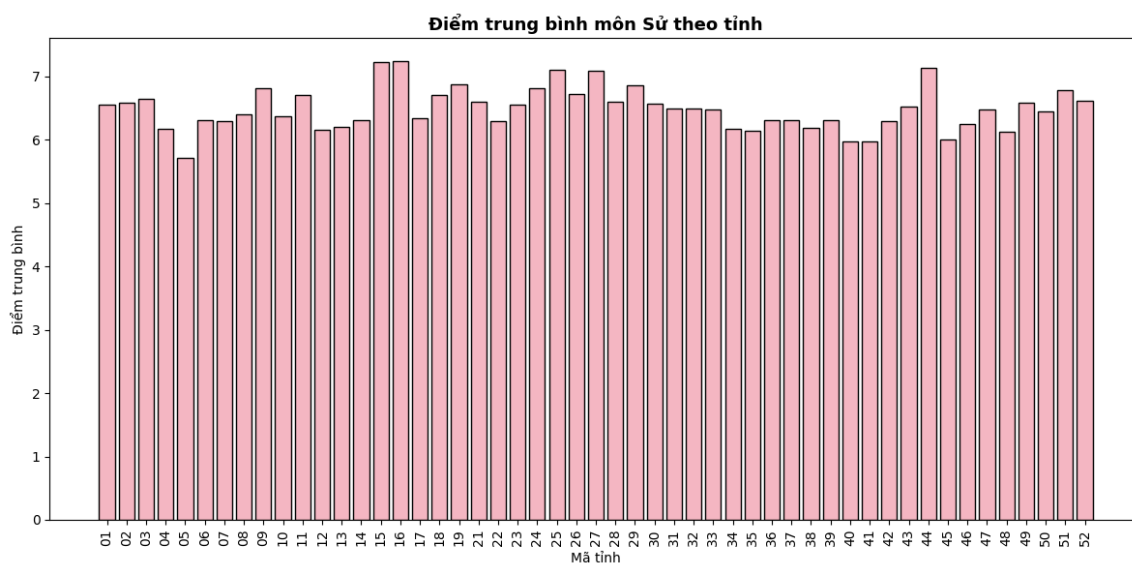


Hình 3: Điểm trung bình môn Sinh theo tỉnh

## Nhận xét:

- Điển hình điểm cao: Thừa Thiên Huế (33), Nghệ An (29), Quảng Ninh (17).
- Điểm thấp rõ rệt: Hà Giang (05), Cao Bằng (06), Tuyên Quang (09).
- Mức dao động rộng hơn môn Ngoại ngữ (khoảng 2 điểm).
- Xu hướng: nhiều tỉnh miền núi phía Bắc có điểm thấp hơn mặt bằng chung.

## Điểm trung bình môn Sử theo tỉnh

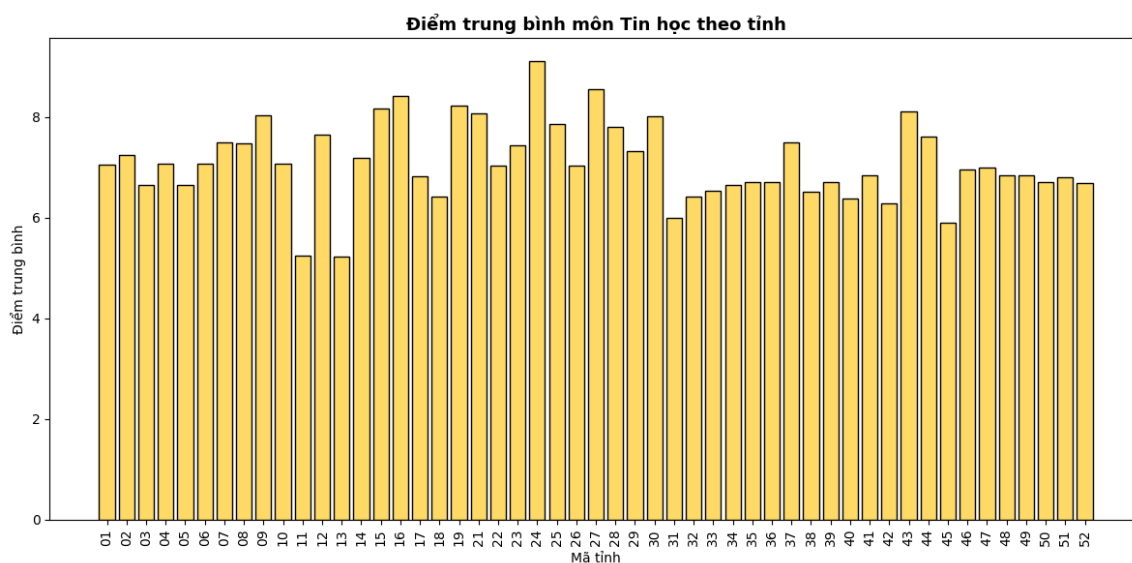


Hình 4: Điểm trung bình môn Sử theo tỉnh

### Nhận xét:

- Các tỉnh dẫn đầu: Nghệ An (29), Thừa Thiên Huế (33), Quảng Ninh (17), Nam Định (25).
- Tỉnh thấp nhất: Điện Biên (62), Hà Giang (05), Cao Bằng (06).
- Môn Sử có chênh lệch khá mạnh giữa các nhóm tỉnh.
- Miền Bắc trung du – miền núi tiếp tục thấp hơn so với đồng bằng và duyên hải.

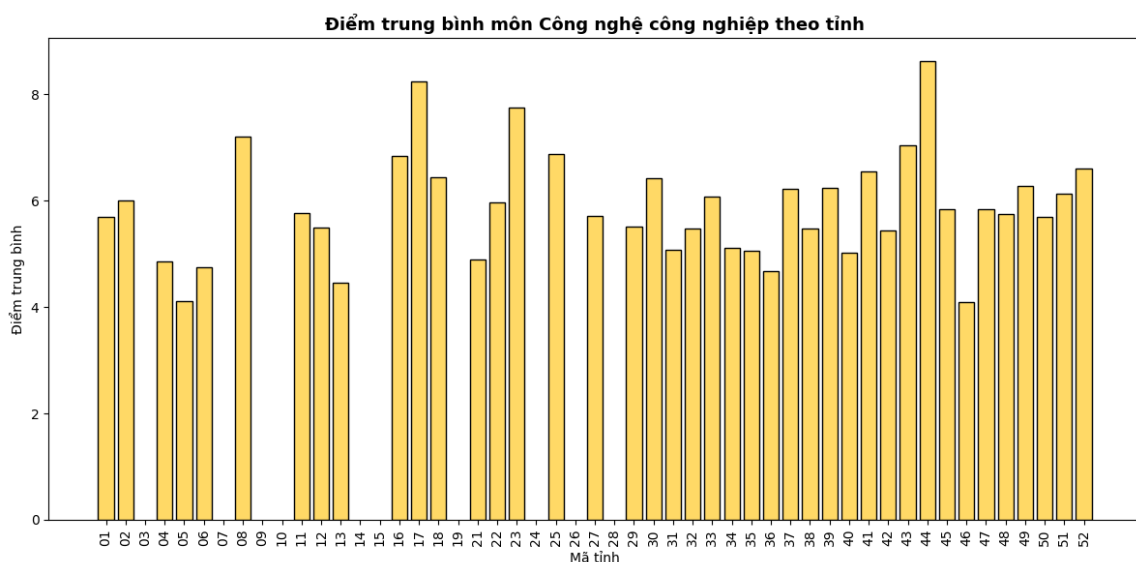
### Điểm trung bình môn Tin học theo tỉnh



Hình 5: Điểm trung bình môn Tin học theo tỉnh

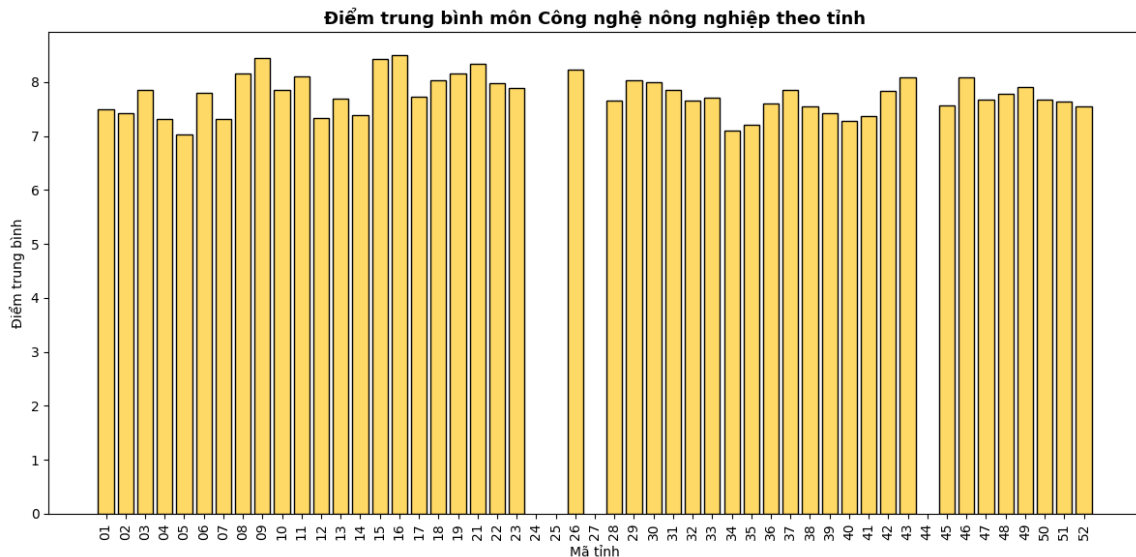
### Nhận xét:

- Điểm rất cao tại: Nam Định (25), Nghệ An (29), Quảng Ninh (17), Khánh Hòa (41).
- Mức thấp nhất thuộc các tỉnh miền núi: Lai Châu (07), Hà Giang (05), Cao Bằng (06).
- Biến thiên lớn nhất trong các môn (dao động gần 3 điểm).
- Sự phân hóa mạnh phản ánh điều kiện hạ tầng công nghệ và chương trình thực hành giữa các vùng.



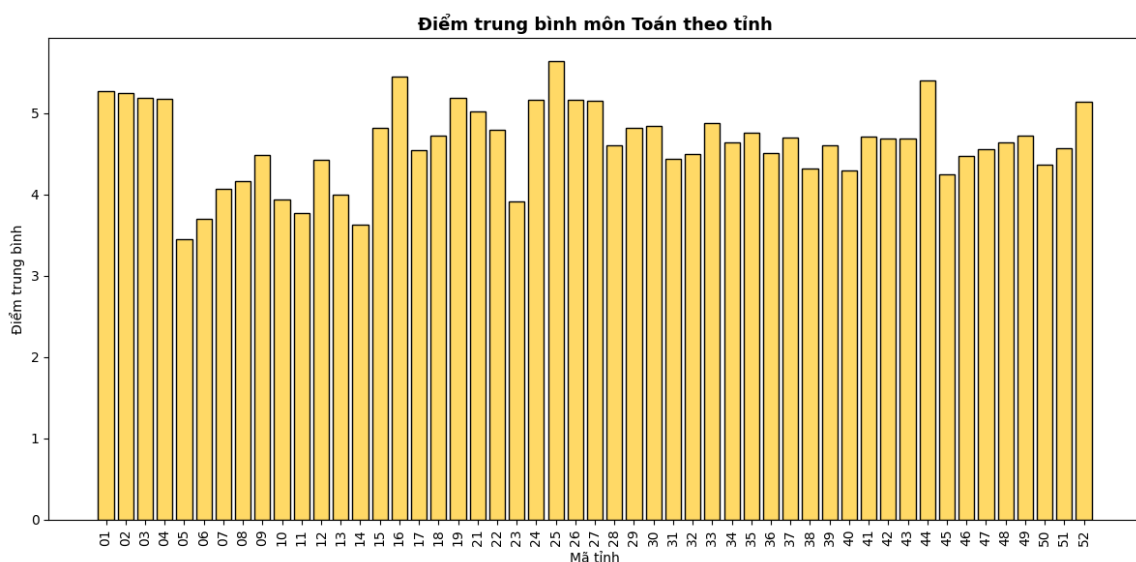
Hình 6: Điểm trung bình môn Công nghệ công nghiệp theo tỉnh

**Nhận xét.** Điểm trung bình phân bố trong khoảng 4.0–8.7, thể hiện sự chênh lệch khá lớn giữa các tỉnh. Những tỉnh đạt điểm cao gồm: Hà Nam (24), Nam Định (25), Quảng Ninh (17), Bình Phước (43). Một số tỉnh có mức điểm thấp hơn: Cao Bằng (06), Sơn La (14), Tây Ninh (46). Sự khác biệt phản ánh mức độ định hướng giáo dục kỹ thuật không đồng đều giữa các địa phương.



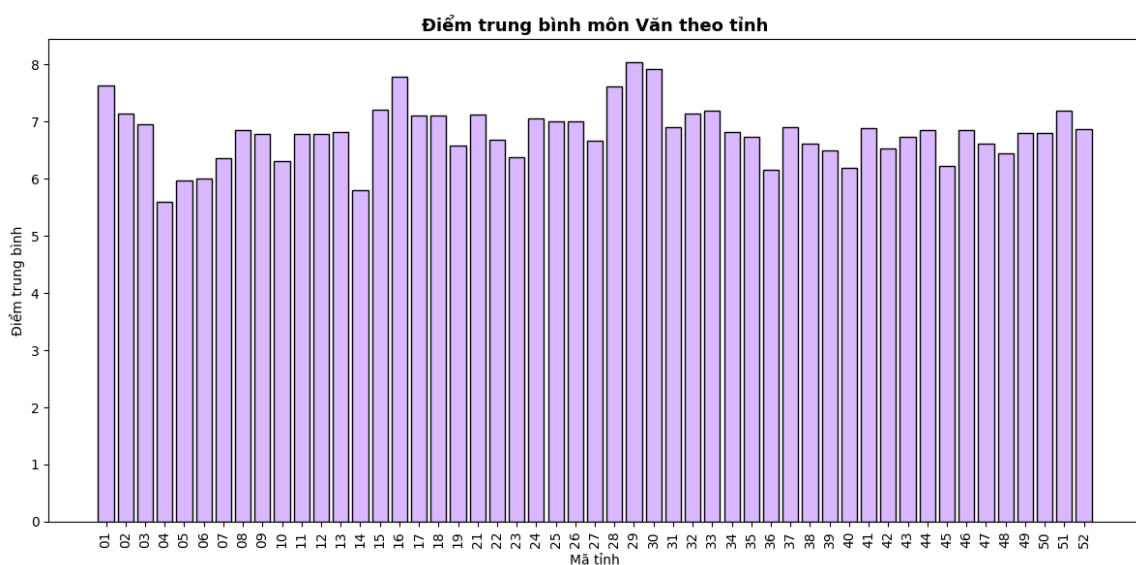
Hình 7: Điểm trung bình môn Công nghệ nông nghiệp theo tỉnh

**Nhận xét.** Mức điểm trung bình khá cao và đồng đều, dao động chủ yếu 7.3–8.5. Các tỉnh dẫn đầu: Hà Nam (24), Nam Định (25), Nghệ An (29). Các tỉnh điểm thấp: Hậu Giang (64), Bạc Liêu (60), Sóc Trăng (59). Đây là môn có độ phân tán nhỏ, cho thấy chất lượng học sinh ổn định trên toàn quốc.



Hình 8: Điểm trung bình môn Toán theo tỉnh

**Nhận xét.** Điểm trung bình môn Toán dao động 3.5–5.5, thấp hơn so với nhiều môn khác. Các tỉnh có mức điểm cao: Hà Nội (01), Nam Định (25), Thái Bình (26), Nghệ An (29). Nhóm tỉnh điểm thấp: Lai Châu (07), Sơn La (14), Hà Giang (05), Kon Tum (36). Chênh lệch gần 2 điểm giữa các tỉnh cho thấy mức độ phân hóa rất rõ theo vùng miền.



Hình 9: Điểm trung bình môn Ngữ văn theo tỉnh

**Nhận xét.** Điểm môn Ngữ văn nằm trong khoảng 6.0–8.0 và khá ổn định giữa các tỉnh. Các tỉnh có điểm cao: Hà Nội (01), Nam Định (25), Nghệ An (29), Thái Bình (26). Nhóm tỉnh thấp hơn: Điện Biên (62), Lai Châu (07), Hà Giang (05). So với Toán, môn Văn thể hiện mức độ chênh lệch nhỏ hơn và có phân bố đồng đều hơn.

### 4.3 Giới thiệu các tổ hợp xét tuyển

Trong phần này, dữ liệu được phân tích dựa trên các tổ hợp xét tuyển phổ biến trong kỳ thi tốt nghiệp THPT. Mỗi tổ hợp được xác định bởi ba môn thành phần, cụ thể như sau:

**A00** : Toán + Vật lý + Hoá học,

**A01** : Toán + Vật lý + Tiếng Anh,

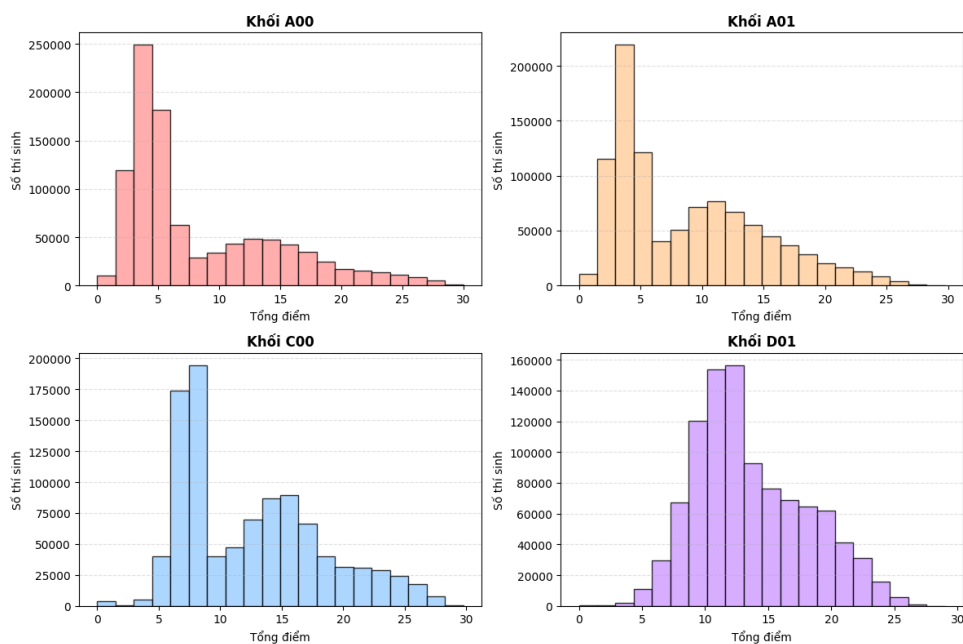
**D01** : Toán + Ngữ văn + Tiếng Anh,

**C00** : Ngữ văn + Lịch sử + Địa lý.

Các tổ hợp trên đại diện cho bốn nhóm định hướng năng lực chính: tự nhiên (A00, A01), ngoại ngữ – tổng hợp (D01), và khoa học xã hội (C00). Phần tiếp theo trình bày phân phối tổng điểm của từng tổ hợp và các insight rút ra từ quan sát dữ liệu.

Mỗi thí sinh được tính tổng điểm theo từng tổ hợp và biểu diễn phân phối bằng biểu đồ histogram.

### 4.4 Biểu đồ phân phối điểm theo tổ hợp



Hình 10: Histogram phân phối tổng điểm theo các tổ hợp A00, A01, C00 và D01

#### 1. Khối C00 có mức điểm cao và ổn định nhất

Phân phối điểm của tổ hợp C00 tập trung mạnh trong khoảng từ 12–20 điểm, hầu như không xuất hiện các giá trị quá thấp. Điều này phản ánh đặc trưng của ba môn thuộc khối C00 (Văn, Sử, Địa), có độ khó trung bình và phân phối điểm tương đối đẹp. Điều này dẫn tới việc các ngành tuyển sinh C00 thường có mức điểm chuẩn ổn định và ít biến động.

## **2. Khối A00 có độ phân tán lớn nhất**

Histogram của A00 cho thấy phân phối điểm trải dài và lệch phải mạnh. Có nhiều thí sinh đạt mức điểm thấp (4–6 điểm), trong khi số lượng thí sinh đạt điểm cao (22–27 điểm) lại rất ít. Điều này phản ánh sự phân hoá rõ rệt năng lực học sinh ở ba môn tự nhiên Toán–Lý–Hoá. Đối với các ngành kỹ thuật, điều này làm cho điểm chuẩn biến động mạnh theo từng năm.

## **3. Khối A01 ổn định hơn A00**

Phân phối A01 đều đặn hơn, ít điểm cực thấp. Sự tham gia của môn Tiếng Anh khiến dải điểm trải rộng và giúp tổng điểm phân bố bớt lệch. Khối A01 phù hợp với những thí sinh vừa có tư duy tự nhiên, vừa có khả năng ngoại ngữ, từ đó làm dải điểm chuẩn các ngành tuyển sinh A01 ổn định hơn so với A00.

## **4. Khối D01 có phân phối “đẹp” nhất**

Histogram của tổ hợp D01 gần như đối xứng, tập trung mạnh tại vùng 12–16 điểm. Sự kết hợp giữa Văn và Anh tạo ra độ trung hoà, giúp phân phối điểm không lệch và ít ngoại lệ. Điều này khiến dự đoán điểm chuẩn cho các ngành xét tuyển D01 trở nên đáng tin cậy hơn.

## **5. Tỷ lệ thí sinh top cao (top 10%) thuộc C00 và D01 lớn hơn**

Dữ liệu cho thấy số lượng thí sinh đạt tổng điểm cao ở các tổ hợp C00 và D01 vượt trội so với A00 và A01. Ngược lại, A00 có tỷ lệ thí sinh top cao rất thấp. Điều này giải thích vì sao các ngành xã hội – nhân văn thường có điểm chuẩn cao, trong khi các ngành kỹ thuật có sự phân tầng mạnh về năng lực.

## **6. Khối A00 mang tính “rủi ro” cao trong xét tuyển**

Sự kéo dài về phía điểm thấp của A00 cho thấy nguy cơ nhiều thí sinh bị rơi khỏi ngưỡng điểm chuẩn nếu một trong ba môn tự nhiên làm bài kém. Điều này làm tăng biến động điểm chuẩn đối với các ngành kỹ thuật.

## **7. Khác biệt năng lực giữa các khối phản ánh xu hướng học tập**

A00 và A01 có tương quan điểm cao do cùng lõi Toán–Lý, trong khi D01 tách biệt rõ và C00 gần như độc lập. Điều này cho thấy tồn tại các nhóm học sinh có thiên hướng năng lực rõ rệt:

- Nhóm tư duy tự nhiên mạnh (A00, A01).
- Nhóm cân bằng, thiên về ngoại ngữ (D01).
- Nhóm thiên về khoa học xã hội (C00).

Việc hiểu rõ cấu trúc này giúp trường đưa ra chiến lược tuyển sinh phù hợp.

## 5 Kiểm tra tương quan giữa các môn

### 5.1 Ma trận tương quan

Trong phần này, tiến hành tính toán hệ số tương quan giữa các môn học để đánh giá mức độ liên hệ về mặt năng lực của thí sinh. Hai nhóm môn được xem xét:

- **Nhóm Tự nhiên:** Toán, Lý, Hoá, Sinh, Tin học, Công nghệ.
- **Nhóm Xã hội:** Văn, Sử, Địa, Giáo dục kinh tế và pháp luật.

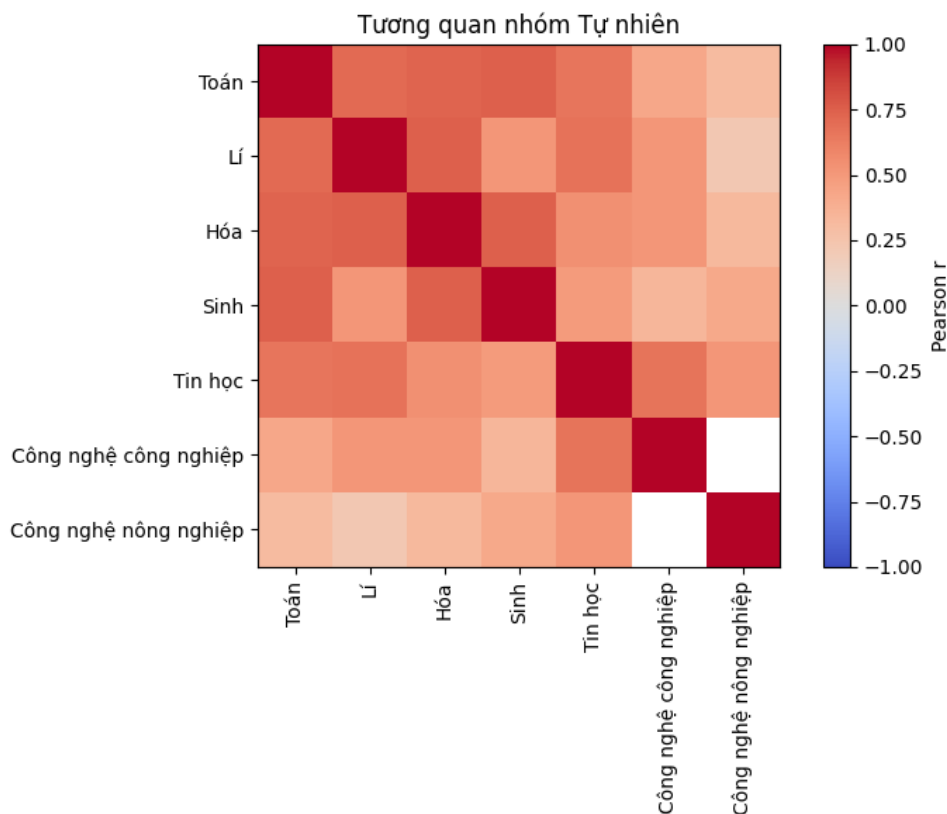
Ngoài ra, các cặp liên nhóm quan trọng cũng được kiểm tra, gồm:

Văn – Ngoại ngữ, Toán – Ngoại ngữ.

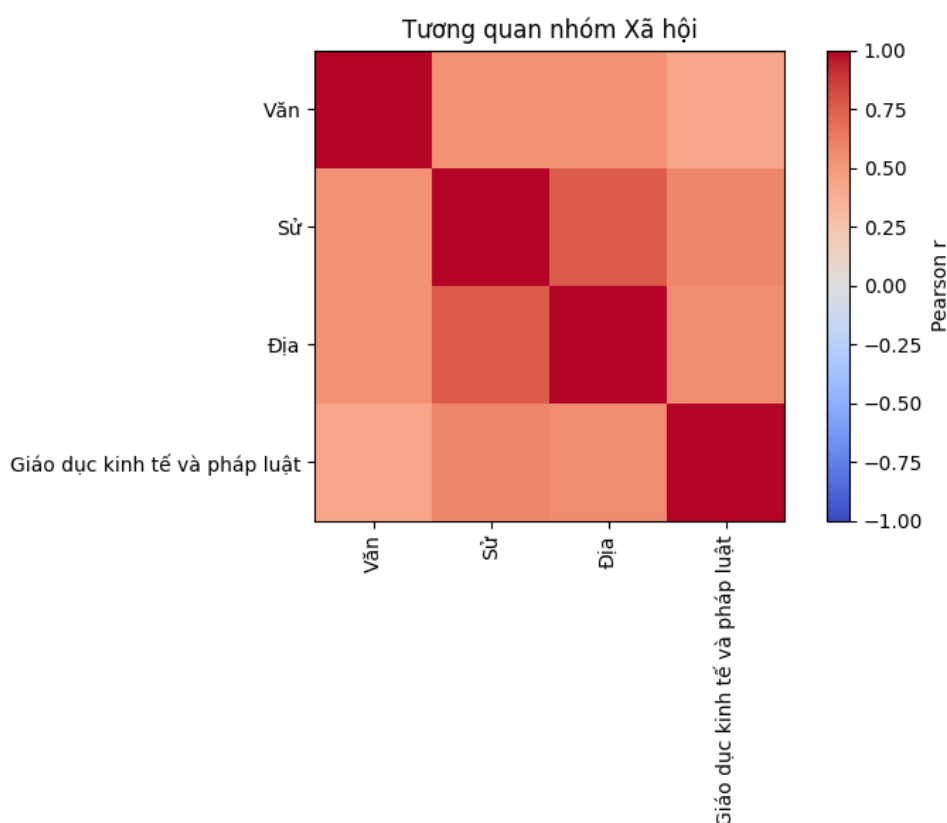
Các ma trận tương quan được xây dựng bằng hệ số **Pearson** vì phù hợp với phân tích mối liên hệ tuyến tính giữa các biến điểm thi. Kết quả được minh hoạ bằng **heatmap**, trong đó:

- Màu đỏ thể hiện tương quan dương mạnh.
- Màu xanh thể hiện tương quan âm hoặc yếu.
- Giá trị trong thang màu nằm ở khoảng  $[-1, 1]$ .

### Hình minh hoạ ma trận tương quan



Hình 11: Ma trận tương quan nhóm Tự nhiên (hệ số Pearson).



Hình 12: Ma trận tương quan nhóm Xã hội (hệ số Pearson).

### 1) Insight nhóm Tự nhiên

- **Bộ ba Toán – Lý – Hoá có tương quan cao nhất** (khoảng 0.60–0.75). Đây là trục năng lực cốt lõi, tạo sự ổn định cho điểm các khối A00/A01.
- **Tin học tương quan khá mạnh với Toán và Lý** (0.40–0.55), phản ánh đặc trưng tư duy logic giữa các môn.
- **Sinh học đóng vai trò cầu nối** giữa nhóm KHTN truyền thống và các môn Công nghệ.
- **Hai môn Công nghệ mang tính độc lập tương đối**: tương quan với Toán–Lý–Hoá thấp ( $< 0.35$ ), cho thấy khác biệt về kỹ năng và nội dung kiểm tra.

### 2) Insight nhóm Xã hội

- **Sử – Địa – GD Kinh tế và pháp luật có tương quan rất cao** (0.55–0.75). Đây là cụm kiến thức xã hội đặc trưng.
- **Môn Văn tách biệt khỏi nhóm Sử–Địa**: tương quan chỉ khoảng 0.35–0.45, phản ánh khác biệt về tư duy ngôn ngữ và ghi nhớ.
- **Văn có tương quan trung bình với Ngoại ngữ**, cho thấy hai môn cùng phản ánh năng lực ngôn ngữ tổng quát.

### 3) Insight liên nhóm

- **Toán có tương quan thấp với Văn/Sử/Địa.** Điều này cho thấy sự tách biệt rõ ràng giữa năng lực phân tích logic và năng lực ngôn ngữ-xã hội.
- **Ngoại ngữ có vai trò “môn cầu nối”:** tương quan trung bình với Văn và thấp-trung bình với Toán.

### 4) Ý nghĩa đối với tuyển sinh

- Khối A00 ổn định do ba môn cốt lõi tương quan mạnh.
- Khối C00 dễ phân tán vì Văn tách biệt khỏi Sử-Địa.
- Điểm Công nghệ không phản ánh năng lực tự nhiên truyền thống.
- Ngoại ngữ là chỉ số tốt để đánh giá năng lực học tập tổng hợp.

Kết luận và đề xuất

## 6 Những nhận định chính

Dựa trên hệ thống biểu đồ và phân tích thống kê đã thực hiện (histogram, density plot, boxplot, phân tích tổ hợp xét tuyển, phân phối theo tỉnh, ma trận tương quan và biểu đồ độ-rớt tốt nghiệp), có thể rút ra một số kết luận quan trọng sau:

### 6.1 Phân phối điểm và mức độ phân hoá

Các biểu đồ phân phối điểm cho thấy sự phân hoá rõ rệt giữa các môn:

- Nhóm môn Tự nhiên (Toán, Lý, Hoá, Sinh) và Ngoại ngữ có phân phối lệch trái, tập trung nhiều ở mức điểm trung bình-thấp, rất ít điểm cao. Đây là nhóm môn dễ kéo giảm điểm xét tuyển và điểm tốt nghiệp.
- Nhóm Xã hội (Sử, Địa, GDCD) có phân phối lệch phải, điểm tập trung nhiều ở mức khá-giỏi, độ phân tán thấp.

### 6.2 Độ biến động và rủi ro theo môn

Biểu đồ boxplot cho thấy:

- Các môn Tự nhiên có độ trải lớn và nhiều điểm ngoại lai thấp, phản ánh mức độ rủi ro cao.
- Môn Văn và Ngoại ngữ có mức biến động mạnh, khó duy trì mức điểm cao đồng đều.
- Nhóm Xã hội ổn định hơn, ít rủi ro điểm thấp.

### 6.3 Tổ hợp xét tuyển và đặc trưng phân phối điểm

Phân tích A00, A01, C00, D01 cho thấy:

- A00 và A01 có dải điểm rộng và phân tán lớn, mức cạnh tranh cao.
- C00 có phân phối ổn định và ít điểm ngoại lai, phù hợp với học sinh thiên hướng phân tích–ghi nhớ.
- D01 cho thấy lợi thế rõ rệt cho thí sinh có năng lực ngôn ngữ tốt.

### 6.4 Khác biệt theo tỉnh và vùng

Biểu đồ điểm trung bình theo tỉnh phản ánh sự khác biệt vùng miền:

- Một số tỉnh/thành phố lớn có điểm trung bình cao và ổn định.
- Các tỉnh vùng khó thường có phân phối điểm thấp hơn, biên độ phân tán lớn.

### 6.5 Tương quan giữa các nhóm môn

Ma trận tương quan Pearson cho thấy:

- Nhóm Tự nhiên (Toán–Lý–Hoá–Sinh) có tương quan cao, thể hiện sự nhất quán về năng lực logic–tính toán.
- Nhóm Xã hội có tương quan trung bình, mỗi môn yêu cầu kỹ năng riêng.
- Ngoại ngữ tương quan vừa với Văn và thấp với các môn Tự nhiên, thể hiện tính độc lập của năng lực ngôn ngữ.

### 6.6 Tỷ lệ đậu – tốt nghiệp

Kết quả phân tích cho thấy:

- Hầu hết thí sinh tốt nghiệp rơi vào một trong hai nhóm: có môn liệt (điểm  $\leq 1$ ) hoặc điểm trung bình tổ hợp thấp.
- Các môn quyết định tỷ lệ tốt bao gồm Toán, Lý và Ngoại ngữ.

## 7 Đề xuất và hướng cải thiện điểm

Dựa trên các insight rút ra từ biểu đồ và dữ liệu, một số giải pháp khả thi gồm:

### 7.1 Tăng cường ôn luyện các môn có rủi ro cao

- Ưu tiên nâng nền tảng ở Toán, Lý và Ngoại ngữ – đây là nhóm môn có tỉ lệ điểm thấp cao nhất.
- Tập trung vào các dạng bài cơ bản để đạt ngưỡng an toàn từ 4–6 điểm trước khi nâng mức.

## **7.2 Lựa chọn tổ hợp xét tuyển theo điểm mạnh**

- A00/A01 phù hợp học sinh có nền tảng Tự nhiên vững.
- C00 phù hợp học sinh mạnh về ghi nhớ và phân tích theo chủ đề.
- D01 nên dành cho thí sinh có khả năng ngôn ngữ tốt.

## **7.3 Ôn tập theo nhóm môn có tương quan**

- Nhóm Toán–Lý–Hoá–Sinh có liên hệ chặt chẽ, nên học theo chuỗi khái niệm để tăng đồng thời nhiều môn.
- Nhóm Sử–Địa–GD&CD nên học theo chủ đề hoặc theo cấu trúc bài để tối ưu thời gian.

## **7.4 Giảm thiểu nguy cơ điểm liệt**

- Thí sinh có nguy cơ điểm liệt cần tập trung luyện đề cơ bản, tránh sai sót nhỏ và ưu tiên mục tiêu đạt từ 2–3 điểm trở lên.
- Đây là yếu tố quan trọng nhất quyết định kết quả tốt nghiệp.

## **7.5 Chiến lược nâng điểm theo mốc mục tiêu**

- Mục tiêu đạt 5 điểm: nắm chắc câu hỏi nhận biết và thông hiểu.
- Mục tiêu 7 điểm: tăng cường dạng vận dụng thấp.
- Mục tiêu 8–9 điểm: luyện tập bộ đề chuẩn hoá, tập trung các chuyên đề khó.

# **8 Tổng kết**

Kết quả phân tích cho thấy sự phân hoá rõ rệt giữa các môn thi, tổ hợp xét tuyển và vùng miền. Việc hiểu đúng đặc trưng phân phối điểm, mối tương quan giữa các môn và nguyên nhân gây rủi ro điểm thấp giúp đề xuất các giải pháp ôn tập hiệu quả hơn. Giữa bối cảnh điểm thi biến động mạnh qua từng năm, việc xây dựng chiến lược học tập hợp lý dựa trên dữ liệu thực tế sẽ góp phần nâng cao kết quả thi và tăng cơ hội xét tuyển cho thí sinh.