



**FPT UNIVERSITY**

# **Computer project – MAS291**

## **Xác suất thông kê**

***Đề tài:*** *Phân tích thống kê về điểm tuyển  
sinh vào 10 của khối chuyên Anh trường  
THPT chuyên Hạ Long.*

***Lớp:*** SE1622

***Giảng viên hướng dẫn :*** Thầy Nguyễn Việt Anh

***Nhóm sinh viên thực hiện :***

Vũ Thùy Linh – HE163140

Nguyễn Ngọc Duy – HE163128

Đình Nhật Duy – HE163347

## Mục Lục

A. Giới thiệu chung về chủ đề.....	2
B. Thu thập dữ liệu và xử lí.....	2
I. Thu nhập và xử lí dữ liệu.....	2
II. Phân tích.....	3
Bài toán 1: Tìm khoảng tin cậy .....	3
Bài toán 2: Kiểm định giả thuyết.....	4
Bài toán 3: Hồi quy tuyến tính .....	6
C. Kết luận.....	7
D. Nguồn tham khảo dữ liệu .....	7

## A. Giới thiệu chung về chủ đề

THPT Chuyên Hạ Long là một trường cấp 3 rất được phụ huynh, học sinh dành sự quan tâm, trong đó lớp chuyên Anh được rất nhiều học sinh tham gia dự tuyển. Khối chuyên Anh của trường sẽ được chia làm 2 lớp: Anh 1 và Anh 2. Vì vậy, phụ huynh và giáo viên thường có những so sánh giữa hai lớp này.

Vậy nên, nhóm em dựa trên điểm tuyển sinh của 2 lớp Anh 1 và Anh 2 niên khóa 2017 - 2020 để tìm hiểu, đánh giá trình độ đầu vào của 2 lớp. Nhóm em đặt ra các vấn đề như sau:

- Trình độ của học sinh lớp Anh 1 có chênh lệch so với lớp Anh 2 không? Cụ thể dựa trên tổng điểm và điểm môn chuyên.
- Tổng điểm thi các môn có liên quan đến điểm chuyên nguyện vọng 1 hay không?

## B. Thu thập dữ liệu và xử lý

### I. Thu nhập và xử lý dữ liệu

Đầu tiên nhóm em có 2 dữ liệu chính: điểm tuyển sinh của tất cả học sinh tham gia kì thi tuyển sinh vào 10 tại địa điểm thi trường THPT Chuyên Hạ Long năm 2017 (1) và danh sách học sinh của 2 lớp niên khóa 2017-2020 (2).

**Dân số:** điểm tuyển sinh vào 10 của lớp Anh 1 và Anh 2 THPT Chuyên Hạ Long của tất cả các khóa.

**Mẫu:** Do không thể có dữ liệu tuyển sinh từ tất cả các khóa, nhóm em lấy mẫu là điểm tuyển sinh vào 10 của lớp Anh 1 và Anh 2 niên khóa 2017-2020 THPT Chuyên Hạ Long.

Nhóm em so sánh và lấy ra được bảng điểm của hai lớp từ danh sách (1) bao gồm điểm của tất cả các khối chuyên. (Cụ thể trong file excel)

Tổng điểm 4 môn (Tổng = Toán+Anh+Văn+Chuyên*2)		
	Anh 1	Anh 2
Average ( $\bar{x}$ )	40.36	39.78
Standard deviation (s)	1.51	1.06
Size (n)	33	33

**Bảng 1.1**

Điểm chuyên (Anh)		
	Anh 1	Anh 2
Average ( $\bar{x}$ )	7.61	7.21
Standard deviation (s)	0.46	0.43
Size (n)	33	33

**Bảng 1.2**

### Giải thích:

- Do điểm trong danh sách (1) sẽ được thống kê theo thứ tự khối chuyên, học sinh tham gia dự tuyển vào lớp chuyên Anh sẽ được xếp riêng và điểm của bài thi môn chuyên Tiếng Anh sẽ được ghi là điểm **NV1**.
- Tổng điểm sẽ được tính theo Tổng = Toán + Anh + Văn + NV1 \* 2

## II. Phân tích

### Bài toán 1: Tìm khoảng tin cậy

Do  $n > 30$ , ta sử dụng công thức:

Large-Sample  
Confidence  
Interval on  
the Mean

When  $n$  is large, the quantity

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

has an approximate standard normal distribution. Consequently,

$$\bar{x} - z_{\alpha/2} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2} \frac{s}{\sqrt{n}} \quad (8-11)$$

is a **large sample confidence interval** for  $\mu$ , with confidence level of approximately  $100(1 - \alpha)\%$ .

$$\alpha = 0.05$$

Khoảng tin cậy 95% của tổng điểm thi trung bình của lớp Anh 1 các khóa là:

$$40.36 - 1.96 \times \frac{1.51}{\sqrt{33}} \leq \mu \leq 40.36 + 1.96 \times \frac{1.51}{\sqrt{33}}$$

$$39.84 \leq \mu \leq 40.87$$

Khoảng tin cậy 95% của tổng điểm thi trung bình của lớp Anh 2 các khóa là:

$$39.78 - 1.96 \times \frac{1.06}{\sqrt{33}} \leq \mu \leq 39.78 + 1.96 \times \frac{1.06}{\sqrt{33}}$$

$$39.42 \leq \mu \leq 40.14$$

Khoảng tin cậy 95% của điểm thi trung bình môn chuyên (NV1) của lớp Anh 1 các khóa là:

$$7.61 - 1.96 \times \frac{0.46}{\sqrt{33}} \leq \mu \leq 7.61 + 1.96 \times \frac{0.46}{\sqrt{33}}$$

$$7.45 \leq \mu \leq 7.77$$

Khoảng tin cậy 95% của điểm thi trung bình môn chuyên (NV1) của lớp Anh 2 các khóa là:

$$7.21 - 1.96 \times \frac{0.43}{\sqrt{33}} \leq \mu \leq 7.21 + 1.96 \times \frac{0.43}{\sqrt{33}}$$

$$7.06 \leq \mu \leq 7.35$$

Nói cách khác,

- 95% tổng điểm trung bình của lớp Anh 1 các khóa nói chung nằm trong khoảng từ 39.84 cho đến 40.87

- 95% tổng điểm trung bình của lớp Anh 2 các khóa nói chung nằm trong khoảng từ 39.42 cho đến 40.14

- 95% điểm môn chuyên trung bình của lớp Anh 1 các khóa nói chung nằm trong khoảng từ 7.45 cho đến 7.77

- 95% điểm môn chuyên trung bình của lớp Anh 2 các khóa nói chung nằm trong khoảng từ 7.06 cho đến 7.35

## **Bài toán 2: Kiểm định giả thuyết**

**Giả thuyết:** Phụ huynh thường cho rằng năng lực của học sinh lớp Anh 1 sẽ cao hơn của học sinh lớp Anh 2.

Nhóm em sẽ kiểm định giả thuyết này bằng cách kiểm định sự giống nhau giữa tổng điểm trung bình của 2 lớp và điểm môn chuyên (NV1) trung bình của 2 lớp.

### **1. Kiểm định về tổng điểm trung bình giữa 2 lớp**

1.1 Dữ liệu ta quan tâm ở đây là  $\mu_1$  và  $\mu_2$ , là hai tổng điểm trung bình lần lượt của 2 lớp Anh 1 và Anh 2 tất cả các khóa.

Giả sử ta có:

1.2.  $H_0: \mu_1 - \mu_2 = 0$  hoặc  $H_0: \mu_1 = \mu_2$ ;

1.3.  $H_1: \mu_1 > \mu_2$ ;

1.4.  $\alpha = 0.05$

1.5. Phương trình thống kê kiểm định là:

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2 - 0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

1.6. Ta sẽ bác bỏ  $H_0$  nếu  $t_0 > t_{0.05,64} = 1.67$

1.7. Tính toán: Từ bảng 1.1 ta có:  $\bar{x}_1 = 40.36$ ,  $\bar{x}_2 = 39.78$ ,  $s_1 = 1.51$ ,  $s_2 = 1.06$ ,  $n_1 = 33$ ,  $n_2 = 33$ . Suy ra:

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2} = \frac{(32)(1.51)^2 + (32)(1.06)^2}{33+33-2} = 1.70$$

$$s_p = \sqrt{1.70} = 1.30$$

và

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2}{1.3 \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{0.58}{1.3 \sqrt{\frac{1}{33} + \frac{1}{33}}} = 1.81$$

1.8. Kết luận: Vì  $t_0 = 1.81 > 1.67$ , ta có thể bác bỏ giả thuyết không. Đó là, ở mức độ ý nghĩa 0.05, ta có bằng chứng rằng tổng điểm trung bình của hai lớp sẽ khác nhau, cụ thể tổng điểm trung bình của lớp Anh 1 cao hơn lớp Anh 2.

## 2. Kiểm định về điểm môn chuyên trung bình giữa 2 lớp

2.1 Dữ liệu ta quan tâm ở đây là  $\mu_1$  và  $\mu_2$ , là hai điểm trung bình của môn chuyên lần lượt của 2 lớp Anh 1 và Anh 2 các khóa nói chung.

Giả sử ta có:

2.2.  $H_0: \mu_1 - \mu_2 = 0$  hoặc  $H_0: \mu_1 = \mu_2$ ;

2.3.  $H_1: \mu_1 > \mu_2$ ;

2.4.  $\alpha = 0.05$

2.5. Phương trình thống kê kiểm định là:

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2 - 0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

2.6. Ta sẽ bác bỏ  $H_0$  nếu  $t_0 > t_{0.05,64} = 1.67$

2.7. Tính toán: Từ bảng 1.2 ta có:  $\bar{x}_1 = 7.61$ ,  $\bar{x}_2 = 7.21$ ,  $s_1 = 0.46$ ,  $s_2 = 0.43$ ,  $n_1 = 33$ ,  $n_2 = 33$ . Suy ra:

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2} = \frac{(32)(0.46)^2 + (32)(0.43)^2}{33+33-2} = 0.198$$

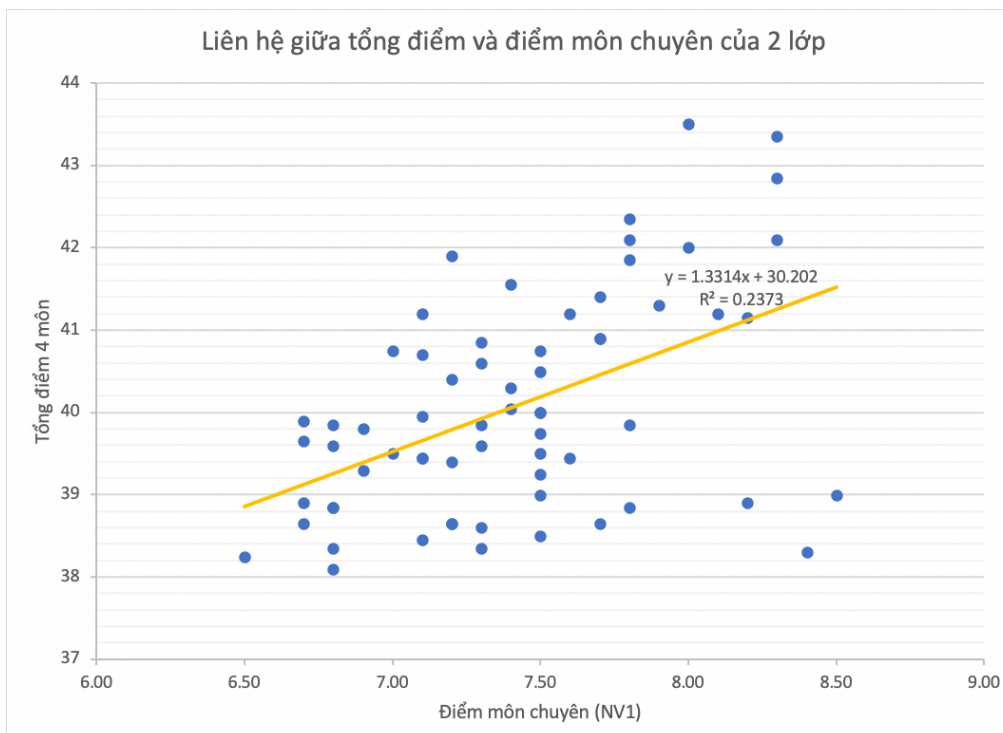
$$s_p = \sqrt{0.198} = 0.44$$

và

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2}{0.44 \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{7.61 - 7.21}{0.44 \sqrt{\frac{1}{33} + \frac{1}{33}}} = 3.69$$

2.8. Kết luận: Vì  $t_0 = 3.69 > 1.67$ , ta có thể bác bỏ giả thuyết không. Đó là, ở mức độ ý nghĩa 0.05, ta có bằng chứng rằng tổng điểm trung bình của hai lớp sẽ khác nhau, cụ thể điểm trung bình môn chuyên của lớp Anh 1 cao hơn lớp Anh 2.

### Bài toán 3: Hồi quy tuyến tính



Ta có đường thẳng hồi quy (regression line) là:

$$y = 1.3314x + 30.202$$

Kiểm định giả thuyết cho regression slope:

$$H_0: \beta_1 = 0 \quad H_1: \beta_1 \neq 0;$$

Ta có  $\widehat{\beta}_1 = 1.3314$ ,  $n = 66$ ,  $\hat{\sigma}^2 = 2.05$  (tính bằng excel)

Phương trình kiểm định là :

$$t_0 = \frac{\widehat{\beta}_1}{se(\widehat{\beta}_1)} = \frac{\widehat{\beta}_1}{\sqrt{\hat{\sigma}^2/S_{xx}}} = \frac{1.3314}{\sqrt{2.05/20.25}} = 4.18$$

$$t_{\alpha/2, n-2} = t_{0.025, 64} = 1.998$$

Do  $|t_0| > t_{\alpha/2, n-2}$  nên ta có bằng chứng để bác bỏ giả thuyết  $H_0 : \beta_1 = 0$

=> Có một mối quan hệ hồi quy tuyến tính giữa điểm thi môn chuyên và tổng điểm các môn của cả hai lớp tuân theo đường thẳng:

$$y = 1.3314x + 30.202$$

## C. Kết luận

Từ những dữ liệu thu được cũng như quá trình phân tích dữ liệu của điểm tuyển sinh hai lớp Anh 1 và Anh 2 niên khóa 2017-2020, ta có thể đi tới những kết luận cho điểm tuyển sinh hai lớp Anh 1 và Anh 2 các khóa nói chung như sau:

- Có sự chênh lệch giữa trình độ đầu vào của Anh 1 và Anh 2, cụ thể điểm tổng và điểm chuyên của lớp Anh 1 cao hơn nhưng sự khác biệt này không quá lớn.
- Có sự liên quan giữa điểm tổng các môn và điểm môn chuyên của 2 lớp. Cụ thể, những học sinh có điểm môn chuyên cao có xu hướng có tổng điểm của các môn cao hơn.

## D. Nguồn tham khảo dữ liệu

Trang web mới và cũ của THPT Chuyên Hạ Long:

<https://c3chuyenhalong.edu.vn/>

<https://c3chuyenhalong.edu.vn/>