



# MAS291 COMPUTER PROJECT

Khảo sát về thời gian tập thể dục trong thời gian học online của sinh viên fpt.

## Members:

HE153663 – Trần Thế Vinh  
HE153664 – Phạm Duy Khánh  
HE153696 – Nguyễn Thành Đạt  
HE153704 – Nguyễn Tiến Đạt

## I. Giới thiệu chung

Dịch COVID-19 hiện tại đang lan tràn tại nhiều tỉnh, thành phố, diễn biến hết sức phức tạp. Việc học online đã và đang được áp dụng tại đa phần các trường học trên cả nước, và Đại học FPT cũng là một trong số đó. Vì vậy, để đảm bảo cho sức khỏe của sinh viên FPT, chúng em đã làm bài khảo sát nhỏ về việc tập thể dục của sinh viên trong thời gian học online để xem dịch bệnh ảnh hưởng đến việc tập thể dục như thế nào.

## II. Thập dữ liệu và xử lý

### 1, Thu thập dữ liệu

- Dữ liệu được nhóm em thu thập từ 64 sinh viên của Đại Học FPT bằng cách khảo sát trực tiếp với biểu mẫu để mọi người điền thông tin thời gian mọi người dành ra để tập thể dục thể thao trong thời gian học online.

#### • Tỷ lệ về giới tính

Giới tính	Tỷ lệ
Nam	65.6%
Nữ	34,4%

#### • Tỷ lệ khu vực về vùng sống

Khu vực	Tỷ lệ
Thành thị	65.6%
Nông thôn	34.4%

#### • Tỷ lệ về ngành học

Ngành học	Tỷ lệ
Công nghệ thông tin	30.2%
Ngôn ngữ	49.2%
Kinh tế	20.6%

#### • Tỷ lệ về thời gian tập/tuần

Thời gian tập trong 1 tuần(buổi)	Tỷ lệ
1 buổi	12.7%
2 buổi	6.3%
3 buổi	25.4%
4 buổi	19%

5 buổi	15.9%
6 buổi	14.3%
7 buổi	6.3%

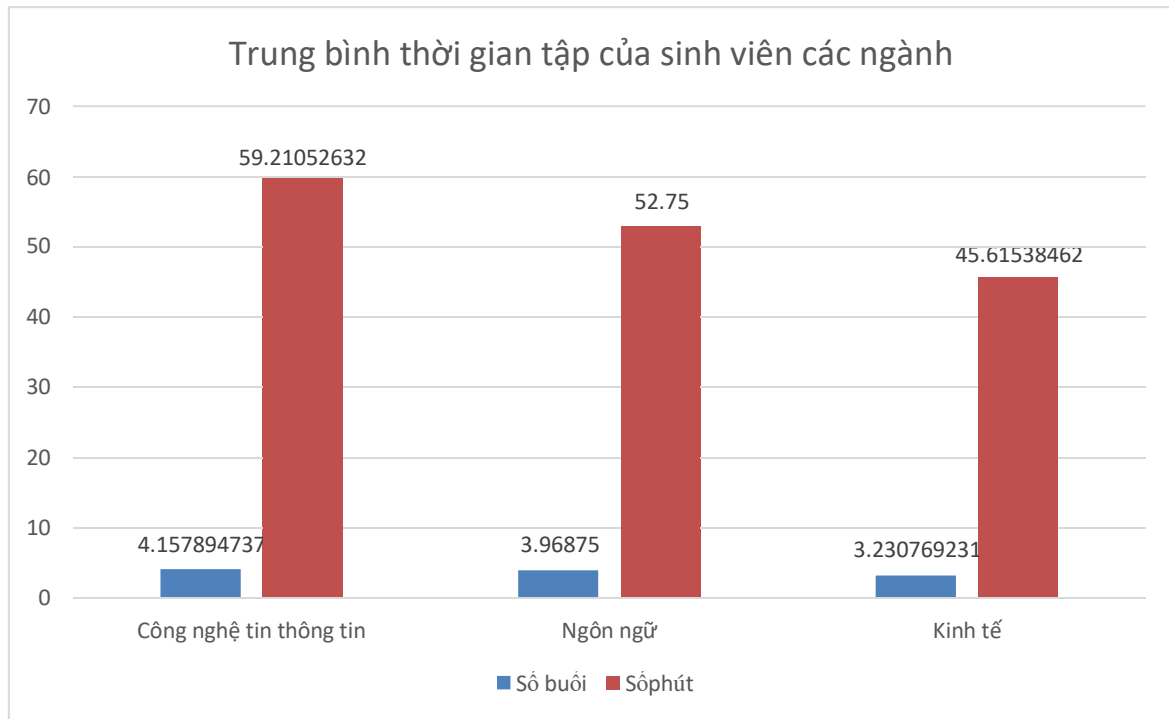
• **Tỉ lệ về thời gian tập/buổi**

Thời gian tập trong 1 buổi(phút)	Số lượng người
10	1
13	1
15	1
17	1
18	1
20	2
23	1
25	2
30	15
60	28
90	8
120	2
130	1

- Sample là toàn bộ sinh viên đã khảo sát có dữ liệu trong file Excel.
- Xử lí dữ liệu:

Paramenter	Trung bình	Độ lệch chuẩn mẫu	Q1	Median	Q3	Max	Min
Thời gian tập/ tuần	3.88	1.71	3	4	5	7	1
Thời gian tập/ buổi	53.22	27.39	30	60	60	130	10

## 2. So sánh tỉ lệ thời gian tập trung bình của Sinh viên các ngành



## 3. Bạn hãy xây dựng khoảng tin cậy 2 phía $\mu$ cho việc hoạt động thể dục thể thao trong thời gian học online : ( $\alpha = 0.05$ )

\*Trả lời:

Áp dụng công thức:

$$\bar{x} - z_{\alpha/2} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2} \frac{s}{\sqrt{n}}$$

Ta có:

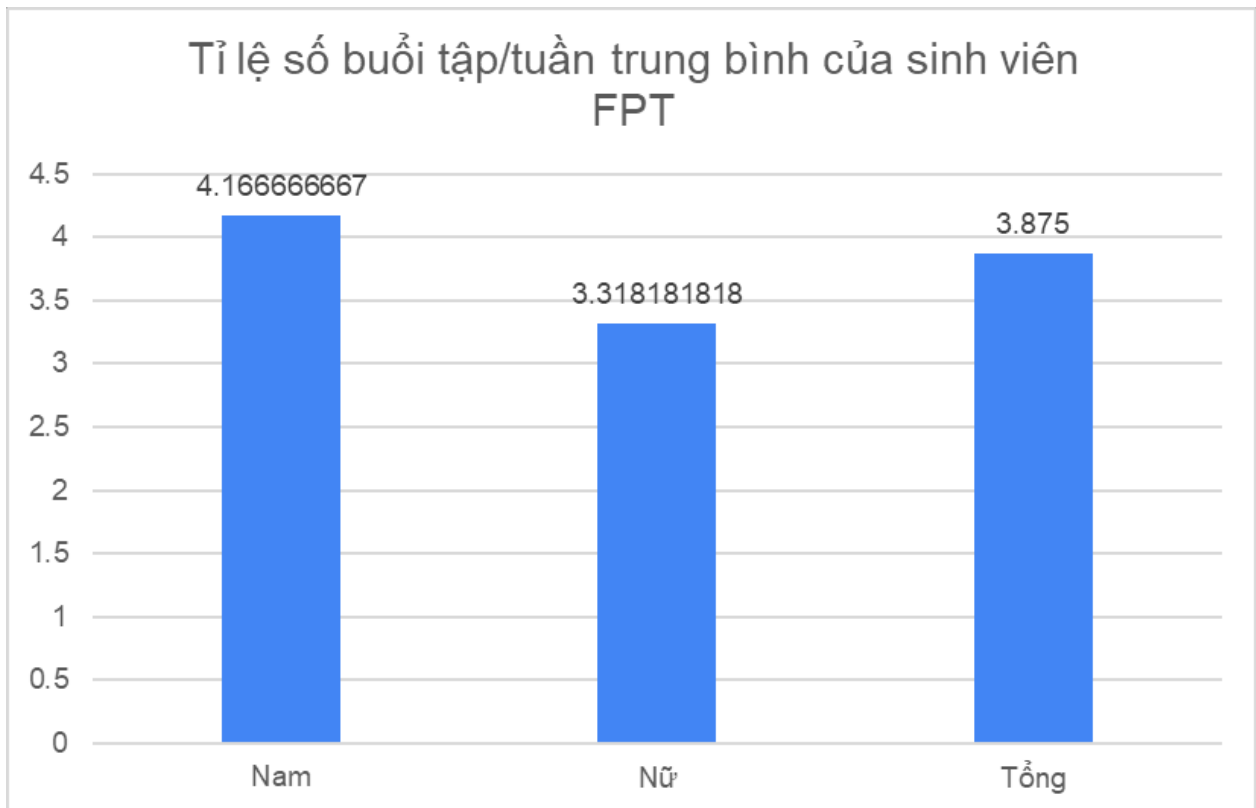
- Khoảng tin cậy 2 phía  $\mu$  của sinh viên cho việc tập luyện số buổi/tuần:

$\bar{x} = 3.88$  ;  $\alpha = 0.05$ ;  $s = 1.71$ ;  $n = 64$ ;  $Z_{\alpha} = 1.96$

$\Rightarrow 3.461 \leq \mu \leq 4.299$

- Khoảng tin cậy 2 phía  $\mu$  của sinh viên cho việc tập luyện số phút/tuần:  
 $\bar{x} = 53.22$ ;  $\alpha = 0.05$ ;  $s = 27.39$ ;  $n = 64$ ;  $Z_{\alpha} = 1.96$   
 $\Rightarrow 46.509 \leq \mu \leq 59.931$

#### 4. So sánh tỉ lệ nam và nữ tập trung bình bao nhiêu buổi / tuần với tỉ lệ sinh viên toàn trường tập trung bình bao nhiêu buổi / tuần



- Từ biểu đồ trên ta có thể suy ra được rằng số buổi tập trung bình / tuần của sinh viên nam trong trường là cao nhất và nữ là thấp nhất, từ đó ta có thể đặt ra :

a. Tin tưởng 95% số buổi tập trung bình/ tuần của sinh viên nữ bằng với tỉ lệ buổi tập trung bình/ tuần của sinh viên toàn trường

- Với  $\alpha = 0.05$ ,  $\bar{X} = 3.3181818$ ,  $\mu_0 = 3.875$ ,  $\sigma = 1.701158251$ ,  $n = 22$

$$H_0: \mu = \mu_0$$

$$H_1: \mu \neq \mu_0$$

Nếu  $H_0$  là đúng thì ta có:  $X \sim N(\mu_x = \mu_0, \sigma_x = \frac{\sigma}{\sqrt{n}})$

$$\text{Bởi vì } P(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}) = 1 - \alpha$$

$$\text{Nên } P(\mu_0 - z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}} < X < \mu_0 + z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

$$\begin{aligned} \rightarrow P(3.875 - 1.96 \times \frac{1.701158251}{\sqrt{22}} < X < 3.875 + 1.96 \times \frac{1.701158251}{\sqrt{22}}) &= 1 - \alpha \\ \rightarrow P(3.164131211 < X < 4.585868789) &= 1 - \alpha \end{aligned}$$

Vì  $X = 3.3181818$  thuộc khoảng trên nên ta không đủ chứng cứ để bác bỏ  $H_0$

b. Tin tưởng 95% số buổi tập trung bình/ tuần của sinh viên nam bằng với tỉ lệ buổi tập trung bình/ tuần của sinh viên toàn trường

- Với  $\alpha = 0.05$ ,  $X = 4.1666667$ ,  $\mu_0 = 3.875$ ,  $\sigma = 1.666260113$ ,  $n = 42$

$$H_0: \mu = \mu_0$$

$$H_1: \mu \neq \mu_0$$

Nếu  $H_0$  là đúng thì ta có:  $X \sim N(\mu_x = \mu_0, \sigma_x = \frac{\sigma}{\sqrt{n}})$

$$\text{Bởi vì } P(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}) = 1 - \alpha$$

$$\text{Nên } P(\mu_0 - z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}} < X < \mu_0 + z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

$$\rightarrow P\left(3.875 - 1.96 \times \frac{1.666260113}{\sqrt{42}} < X < 3.875 + 1.96 \times \frac{1.666260113}{\sqrt{42}}\right) = 1 - \alpha$$

$$\rightarrow P(3.371065346 < X < 4.378934654) = 1 - \alpha$$

Vì  $X = 4.1666667$  thuộc khoảng trên nên ta không đủ chứng cứ để bác bỏ  $H_0$

c. Thử xem tỉ lệ sinh viên nữ dành cho số buổi tập trung bình / tuần lớn hơn tỉ lệ toàn bộ sinh viên dành cho số buổi tập trung bình / tuần không vượt quá 0.5

$$H_0: p = p_0$$

$$H_1: p < p_0$$

$$z_o = \frac{x - np_0}{\sqrt{np_0(1-p_0)}} \sim N(0, 1) \text{ với } p_0 = 0.5, x = 7, n = 22$$

$$\text{Bác bỏ } H_0: p = 0.5 \text{ if } z_o < -z_{0.05} = -1.645$$

$$\text{Tính } z_o = \frac{7 - 22 \times 0.5}{\sqrt{22 \times 0.5 (1-0.5)}} = -1.705605731$$

$\rightarrow$  Ta không đủ chứng cứ để bác bỏ  $H_0$

d. Thử xem tỉ lệ sinh viên nam dành cho số buổi tập trung bình / tuần lớn hơn tỉ lệ toàn bộ sinh viên dành cho số buổi tập trung bình / tuần không vượt quá 0.5

$$H_0: p = p_0$$

$$H_1: p < p_0$$

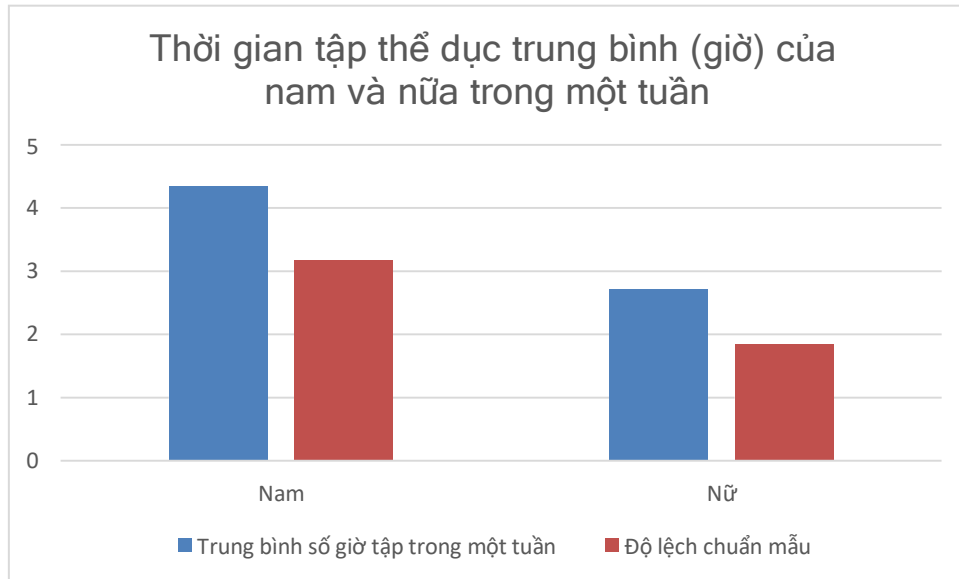
$$z_o = \frac{x - np_0}{\sqrt{np_0(1-p_0)}} \sim N(0, 1) \text{ với } p_0 = 0.5, x = 29, n = 42$$

$$\text{Bác bỏ } H_0: p = 0.5 \text{ if } z_o > z_{0.05} = 1.645$$

$$\text{Tính } z_o = \frac{29 - 42 \times 0.5}{\sqrt{42 \times 0.5 (1-0.5)}} = 2.468853599$$

$\rightarrow$  Có thể bác bỏ  $H_0$

## 5. So sánh tỉ lệ tập thể dục trung bình giữa nam và nữ



Nam:  $\bar{X}_1 \approx 4.34$        $S_1 \approx 3.18$        $n_1 = 42$

Nữ:  $\bar{X}_2 \approx 2.72$        $S_2 \approx 1.85$        $n_2 = 22$

a) Giả sử  $\sigma_1^2 = \sigma_2^2$

$H_0: \mu_1 - \mu_2 = 0$

$H_1: \mu_1 \neq \mu_2$

(\*) Xét giả định  $H_0$ , ta có:

Bậc tự do:  $D_f = n_1 + n_2 - 2 = 62$

95% là khoảng tin cậy  $\rightarrow \alpha = 0.05$

$t_{\alpha/2, D_f} \approx 2.00$

Có:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$\rightarrow S_p^2 \approx 8.17 \rightarrow S_p \approx 2.86$



$$H_0: \mu_1 - \mu_2 = \Delta_0$$

$$T_0 = \frac{\bar{X}_1 - \bar{X}_2 - \Delta_0}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\Delta_0 = 0$$

$$\rightarrow t_0 \approx 2.15$$

Vậy  $t_0 > t_{0.025, 62} \Rightarrow$  Loại bỏ giả định  $H_0$

b) Giả sử  $\sigma_1^2 \neq \sigma_2^2$

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_1: \mu_1 \neq \mu_2$$

(\*) Xét giả định  $H_0$ , ta có:

$$v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{n_1 - 1} + \frac{(S_2^2/n_2)^2}{n_2 - 1}}$$

$\rightarrow$  Bậc tự do:  $D_f \approx 61,21$

95% là khoảng tin cậy  $\rightarrow \alpha = 0.05$

$$\rightarrow t_{\alpha/2, D_f} \approx 2,00$$

$$\Delta_0 = 0$$

Có:

$$T_0^* = \frac{\bar{X}_1 - \bar{X}_2 - \Delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$$\Rightarrow t_0 \approx 2,57$$

Vậy  $t_0 > t_{0.025, 61.21} \Rightarrow$  Loại bỏ giả định  $H_0$

## 6. Tiến hành kiểm định giả thiết cho Regression Slope

- Với x là thời gian tập/ tuần.
- Y là thời gian tập/buổi có:

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	26.32053	7.716531	3.410927	0.001164	10.88517	41.75589	10.88517	41.75589
X Variable 1	7.13024	1.831152	3.893855	0.000251	3.467391	10.79309	3.467391	10.79309

Từ bảng trên, ta có:

$$B_1 = 7.13; \text{ se}(B_1) = 1.83$$

Kiểm định giả thiết đối với linear regression:

$$H_0: B_1=0$$

$$H_1: B_1 \neq 0$$

$$n = 60; \alpha = 0.01$$

Áp dụng công thức:

$$t_0 = \frac{\hat{\beta}_1}{\sqrt{\hat{\sigma}^2/S_{xx}}} = \frac{\hat{\beta}_1}{\text{se}(\hat{\beta}_1)}$$

$$\Rightarrow t_0 = 7.13/1.83 = 3.90$$

Vì giá trị tham chiếu là  $t_{0.005, 60} = 2,66 \Rightarrow$  giá trị của  $t_0$  thống kê thử nghiệm nằm rất xa vùng tới hạn ngụ ý rằng có đủ chứng cứ để bác bỏ  $H_0 : B_1 = 0$

Thời gian tập/tuần có quan hệ với thời gian tập /buổi.

### **III. Kết luận**

Từ những dữ liệu thu thập cộng với quá trình phân tích, có thể kết luận như sau:

- Tỷ lệ về thời gian tập của sinh viên có sự chênh lệch lớn giữa nam và nữ, thành thị và nông thôn.
- Thời gian tập/tuần có quan hệ với thời gian tập/buổi.
- Các ngành học không ảnh hưởng nhiều đến tỷ lệ tập thể dục của sinh viên.
- Tỷ lệ về thời gian tập trung bình trong tuần có sự biến động đáng kể .