

# Khảo sát sinh hoạt một ngày của sinh viên trong mùa dịch

Nhóm 1 – Lớp SE1606

HE153729	Đinh Văn Tiến
HE153107	Nguyễn Lê Hùng
HE153099	Nguyễn Quang Linh
HE153101	Vũ Ngọc Sơn

## I, Tổng quan

Ở mỗi lứa tuổi khác nhau đều có những đặc điểm tâm và chịu sự chi phối từ những yếu tố khác nhau. Trong những lứa tuổi, sinh viên là một nhóm xã hội đặc biệt, khi không có gia đình bên cạnh sẽ giúp thói quen sinh hoạt của sinh viên sẽ có tính độc lập nhiều hơn. Vì thế, chúng em đã khảo sát những thói quen của sinh viên khi ở nhà, để xem đại dịch đã tác động đến sinh hoạt hằng ngày của sinh viên như thế nào?

## II, Thu thập dữ liệu

- Dữ liệu được thu thập ngẫu nhiên từ 100 sinh viên trong trường Đại Học FPT bằng cách khảo sát trực tiếp từ mỗi bạn sinh viên, chúng em tạo một form cho mọi người có thể điền thông tin là thời gian học, thời gian ngủ, thời gian chơi và thời gian cho các hoạt động khác.

(<https://docs.google.com/forms/d/e/1FAIpQLSf4ustBkHYbu-CjmG6ec6tIgM1YKhoyJThnZEDbJC-GqkKfpQ/viewform>)

- Population ở đây là toàn bộ sinh viên của trường FPT.
- Sample sẽ được lấy từ toàn bộ sinh viên đã qua khảo sát trong file excel và được xử lý dữ liệu như bảng dưới đây.

Paramenter	Sample standard						
	Average	deviation	Q1	Median	Q3	Min	Max
Learn	5.55	2.45515331	4	6	7	0	12
Play	4.63	2.568869578	2	5	6	0	12
Sleep	9.18	1.816757017	8	10	10	5	14
Other	4.64	2.32039704	3	4	6	0	10
Ability	2.2	0.942809042	2	2	3	0	4

Với các công thức

If the  $n$  observations in a sample are denoted by  $x_1, x_2, \dots, x_n$ , the **sample mean** is

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n} \quad (6-1)$$

If  $x_1, x_2, \dots, x_n$  is a sample of  $n$  observations, the **sample variance** is

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad (6-3)$$

The **sample standard deviation**,  $s$ , is the positive square root of the sample variance.

The **first or lower quartile**,  $q_1$ , is a value that has approximately one-fourth (25%) of the observations below it and approximately 75% of the observations above.

The **second quartile**,  $q_2$ , has approximately one-half (50%) of the observations below its value. The second quartile is *exactly* equal to the **median**.

The **third or upper quartile**,  $q_3$ , has approximately three-fourths (75%) of the observations below its value. As in the case of the median, the quartiles may not be unique.

Và khi dùng trong excel thì có các câu lệnh như sau:

Average =AVERAGE(:)

Standard deviation =STDEV.S(:)

Q1 =QUARTILE(:,1)

Median =QUARTILE(:,2)

Q3 =QUARTILE(:,3)

Max =MAX()

Min =MIN()

### III, Câu hỏi

1. Hãy xây dựng khoảng tin cậy  $\mu$  của việc sinh viên dành thời gian cho việc học, chơi và ngủ? (Với  $\alpha = 0.05$ )

Công thức tính khoảng tin cậy  $\mu$ :

When  $n$  is large, the quantity

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

has an approximate standard normal distribution. Consequently,

$$\bar{x} - z_{\alpha/2} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2} \frac{s}{\sqrt{n}} \quad (8-13)$$

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

- Khoảng tin cậy  $\mu$  của sinh viên dành cho việc học:

Với  $x = 5.55$ ,  $\alpha = 0.05$ ,  $s = 2.45515331$ ,  $n = 100$

Thì khoảng tin cậy là  $5.6879 \leq \mu \leq 6.03121$

- Khoảng tin cậy  $\mu$  của sinh viên dành cho việc giải trí:

Với  $x = 4.63$ ,  $\alpha = 0.05$ ,  $s = 2.26887$ ,  $n = 100$

Thì khoảng tin cậy là  $4.1853 \leq \mu \leq 5.0747$

- Khoảng tin cậy  $\mu$  của sinh viên dành cho việc ngủ:

Với  $x = 9.18$ ,  $\alpha = 0.05$ ,  $s = 1.816757$ ,  $n = 100$

Thì khoảng tin cậy là  $8.824 \leq \mu \leq 9.536$

2, So sánh tỉ lệ sinh viên K14 và K15 dành cho việc học với tỉ lệ sinh viên toàn trường dành cho việc học.



Từ biểu đồ ta có thể suy luận được giờ học của sinh viên k14 là cao nhất và k15 là thấp nhất, câu hỏi đặt ra là:

a, Tin tưởng 95% tỉ lệ số thời gian học của sinh viên K15 bằng với tỉ lệ học của toàn trường FPT, với  $\alpha = 0.05$ ,  $X = 5.157894737$ ,  $\mu_0 = 5.55$ ,  $\sigma = 2.87097217$ ,  $n = 57$

$$H_0: \mu = \mu_0$$

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

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

$$\Rightarrow Z = \frac{X - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

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

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

$$\Rightarrow P(4.804671675 < X < 6.295328) = 1 - \alpha$$

Vì  $X = 5.157894737$  thuộc khoảng trên nên ta không đủ chứng cứ để bác bỏ  $H_0$

b, Tin tưởng 95% tỉ lệ số thời gian học của sinh viên K14 bằng với tỉ lệ học của toàn trường FPT, với  $\alpha = 0.05$ ,  $X = 6.413793103$ ,  $\mu_0 = 5.55$ ,  $\sigma = 1.323341047$ ,  $n = 29$

$$H_0: \mu = \mu_0$$

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

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

$$\Rightarrow Z = \frac{X - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

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

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

$$\Rightarrow P(5.06835 < X < 6.031647) = 1 - \alpha$$

Vì  $X = 6.413793103$  không thuộc khoảng trên nên đủ chứng cứ để bác bỏ  $H_0$

c, Kiểm định giả thiết tỉ lệ sinh viên k14 dành cho việc học lớn hơn tỉ lệ toàn bộ trường dành cho việc học không vượt quá 0.5

$$H_0: p = p_0$$

$$H_1: p < p_0$$

$$\text{Với } \alpha = 0.05$$

$$z_o = \frac{x - np_0}{\sqrt{np_0(1-p_0)}} \sim N(0, 1) \text{ với } p_0 = 0.5, x = 25, n = 57$$

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

$$\text{Tính } z_o = -0.9272$$

Suy ra được ta không đủ chứng cứ để bác bỏ  $H_0$

d, Kiểm định giả thiết tỉ lệ sinh viên k14 dành cho việc học lớn hơn tỉ lệ toàn bộ trường dành cho việc học không vượt quá 0.5

$$H_0: p = p_0$$

$$H_1: p < p_0$$

$$\text{Với } \alpha = 0.05$$

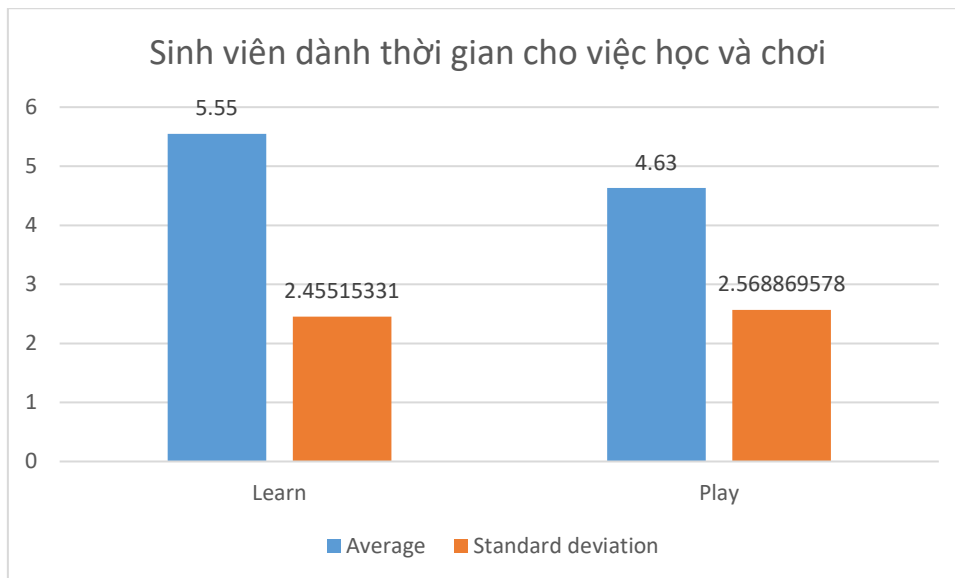
$$z_o = \frac{x - np_0}{\sqrt{np_0(1-p_0)}} \sim N(0, 1) \text{ với } p_0 = 0.5, x = 20, n = 29$$

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

$$\text{Tính } z_o = 2.042$$

Suy ra được ta không đủ chứng cứ để bác bỏ  $H_0$

3, Có 2 tập dữ liệu về sinh viên dành thời gian học và sinh viên dành thời gian chơi. Kiểm định chúng có khác nhau hay không? Với  $\alpha = 0.05$



a, Giả định variance bằng nhau

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

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

$$\alpha = 0.05$$

$$t_0 = \frac{x_1 - x_2 - 0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Bác bỏ  $H_0$  nếu  $t_0 > t_{0.025, 198} = 1.97$  hoặc  $t_0 < -t_{0.025, 198} = -1.97$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = 6.313$$

$$s_p = 2.512655$$

$$\Rightarrow t_0 = 2.589$$

Kết luận: Chúng ta đã đủ chứng cứ để bác bỏ  $H_0$

b, Giả định variance khác nhau

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

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

$$\alpha = 0.05$$

$$t_0 = \frac{x_1 - x_2 - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

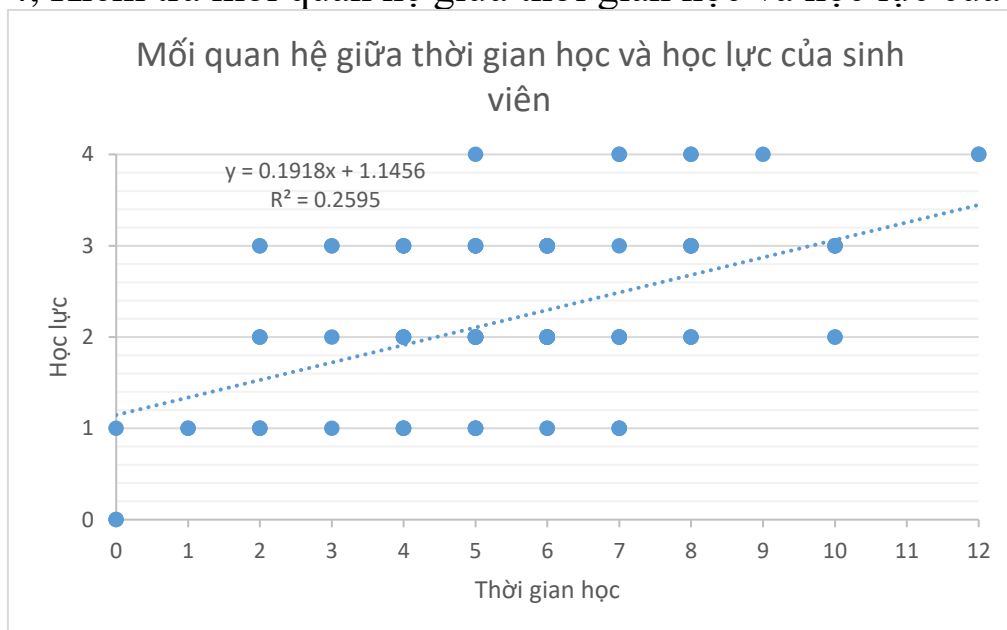
$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_s^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} = 197$$

Bác bỏ  $H_0$  nếu  $t_0 > t_{0.025,197} = 1.97$  hoặc  $t_0 < -t_{0.025,197} = -1.97$

$$\Rightarrow t_0 = 2.589$$

Kết luận: Chúng ta đã đủ chứng cứ để bác bỏ  $H_0$

#### 4, Kiểm tra mối quan hệ giữa thời gian học và học lực của sinh viên



Kiểm định giả thiết đối với linear regression:

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

Với  $\alpha = 0.05$ ,  $\hat{\beta}_1 = 0.1918$ ,  $n = 100$ ,  $se(\hat{\beta}_1) = 0.032728$

$$\Rightarrow t_0 = \frac{\hat{\beta}_1}{\frac{\sqrt{\sigma^2}}{\sqrt{S_{xx}}}} = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = 5.86$$

Với  $t_{0.025,98} = 1.984$  thì giá trị của  $t_0$  nằm ngoài khoảng

$\Rightarrow$  Có đủ chứng cứ để bác bỏ  $H_0: \beta_1 = 0$

Có thể nói rằng việc sinh viên dành thời gian cho học thì có liên quan tới học lực của họ.

#### IV, Kết luận

Từ những dữ liệu thu được cũng như quá trình phân tích dữ liệu, ta có thể đi tới những kết luận như sau :

Có thể thấy rằng đại dịch có những tác động tới những thói quen như việc học tập cũng như vui chơi của các sinh viên.

Ở độ tuổi khác nhau, có thể thấy sự khác biệt cho việc dành thời gian cho việc học ( cụ thể là giữa sinh viên k14 với các sinh viên khác cũng như sinh viên toàn trường ).

Dưới tác động của đại dịch ảnh hưởng thì nhìn chung sinh viên dành thời gian cho việc học vẫn cao hơn so với vui chơi tuy nhiên sự khác biệt là không cao.