

Project MAS291

I. Members

1. Nguyễn Đức Hải– HE150561
2. Nguyễn Văn Tuyên– HE150819
3. Trần Đức Tuấn– HE151203

II. Data

Class A

No	Mark	No	Mark
1	9	16	10
2	8.5	17	7.7
3	7.5	18	8
4	7.6	19	9.7
5	9.8	20	10
6	7.6	21	7.8
7	8	22	7.5
8	7	23	8.5
9	5.5	24	6.5
10	9	25	9.5
11	9.8	26	7.5
12	9.2	27	4.8

13	6.5	28	10
14	4.5	29	7.5
15	7	30	7.5

Class B

No	Mark	No	Mark
1	6.6	16	10
2	6	17	0
3	5.5	18	8.2
4	6	19	6.6
5	4.2	20	10
6	6.2	21	6.6
7	4.4	22	7
8	8.9	23	0
9	8.1	24	6.6
10	6.9	25	9.2
11	5.6	26	8.8
12	9	27	4.4
13	1.6	28	5.2
14	5.2	29	3.6
15	7	30	6.2

III. Analysis

Section 1: The mean and standard deviation of the two groups

- Group A mean and standard deviation:

- $\sum x^2 = 1968.5$
- $(\sum x)^2 = 239^2$
- $n = 30$

$$\bar{x} \approx 7.97$$

$$S = \sqrt{\frac{n \sum x^2 - (\sum x)^2}{n(n-1)}} \approx \sqrt{\frac{30 \sum x^2 - (\sum x)^2}{30(30-1)}} \approx 1.49$$

- Group B mean and standard deviation:

- $\sum x^2 = 1311.44$
- $(\sum x)^2 = 183.6^2$
- $n = 30$

$$\bar{x} \approx 6,12$$

$$S = \sqrt{\frac{n \sum x^2 - (\sum x)^2}{n(n-1)}} \approx \sqrt{\frac{30 \sum x^2 - (\sum x)^2}{30(30-1)}} \approx 2.545$$

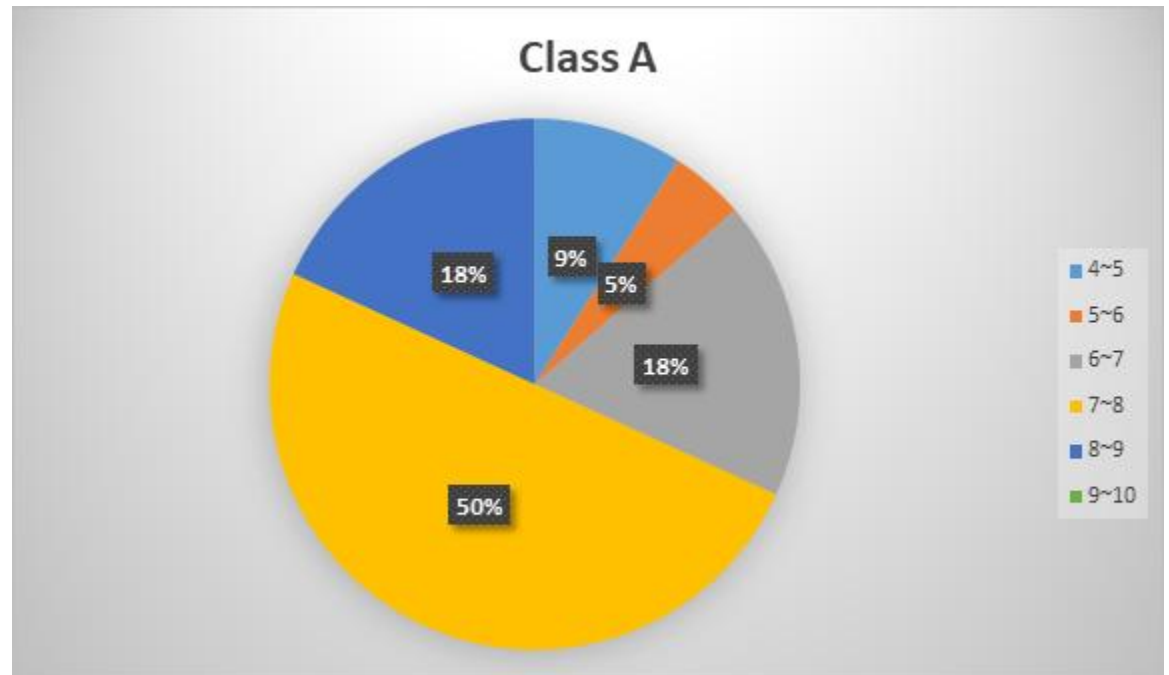
Section 2. Build frequency distribution and histogram for test data of 2 groups

2.1. Frequency distribution (Phân bố tần số)

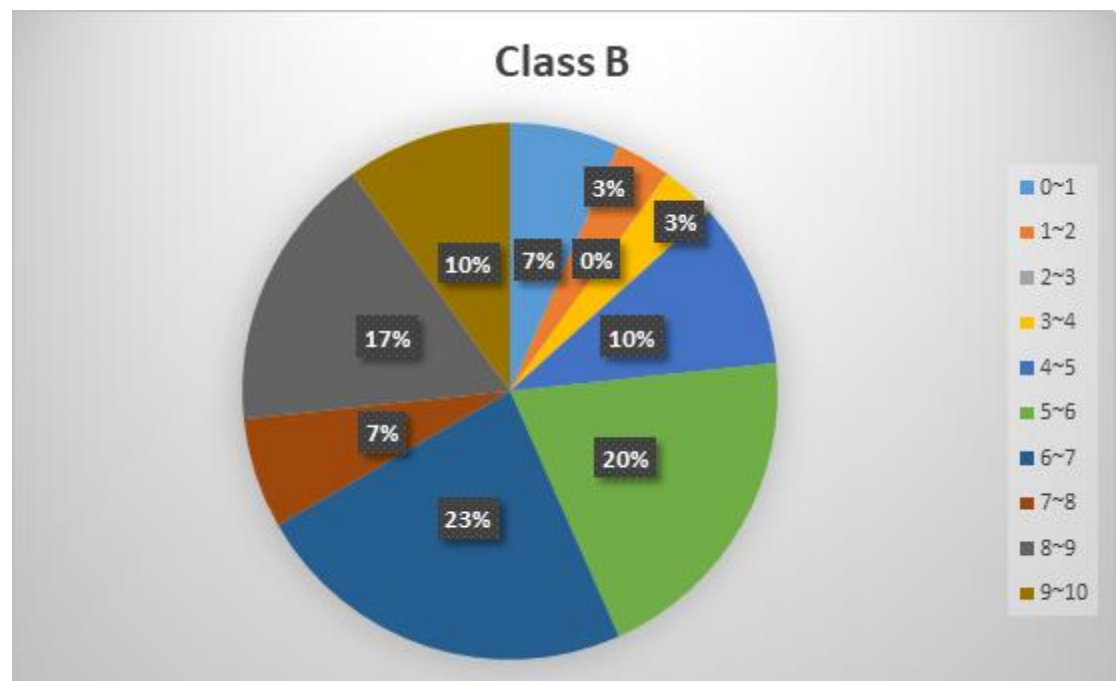
Mark	0 ~ 1	1 ~ 2	2 ~ 3	3 ~ 4	4 ~ 5	5 ~ 6	6 ~ 7	7 ~ 8	8 ~ 9	9 ~ 10
Frequency of Group A	0	0	0	0	2	1	4	11	4	8
Frequency of Group B	2	1	0	1	3	6	7	2	5	3

2.2. Histogram (Biểu đồ)

- A class chart:



- B class chart:



Section 3. Xây dựng Khoảng tin cậy hai phía 95% cho điểm thi trung bình và điểm thi trung bình có điểm thi > 5. (của hai lớp)

According to:

$$\bar{x} - t_{\alpha/2, n-1} s/\sqrt{n} \leq \mu \leq \bar{x} + t_{\alpha/2, n-1} s/\sqrt{n}$$

- Group A:

$$n = 30 \Rightarrow n-1 = 29 \Rightarrow t_{0,025;29} = 2,05$$

$$s = 1.49 \Rightarrow t_{0,025;29} * s/\sqrt{n} = 2,05 * 1.49/\sqrt{30} = 0.557$$

$$\Rightarrow 7.97 - 0.557 \leq \mu \leq 7.97 + 0.557$$

$$\Rightarrow 7.233 \leq \mu \leq 8.347$$

- Group B:

$$n = 30 \Rightarrow n-1 = 29 \Rightarrow t_{0,025;29} = 2,05$$

$$s = 2.545 \Rightarrow t_{0,025;29} * s/\sqrt{n} = 2,05 * 2.545/\sqrt{30} = 0,953$$

$$\Rightarrow 6,12 - 0,953 \leq \mu \leq 6,12 + 0,953$$

$$\Rightarrow 5,167 \leq \mu \leq 7,073$$

Section 4.

-Nếu chúng ta chỉ lấy những sinh viên đạt số điểm > 5 :

- Class A:

Sample mean: $\bar{x} = 8.2$

Sample size: $n = 28$

- $\sum x^2 = 1925.21$

- $(\sum x)^2 = 229.7^2$
- $n = 28$

Standard Deviation; $s = \sqrt{\frac{n \sum x^2 - (\sum x)^2}{n(n-1)}} \approx \sqrt{\frac{28 * 1925.21 - 229.7^2}{28(28-1)}} \approx 1.23$

- Class B:

Sample mean: $\bar{x} = 7.191$

Sample size: $n = 23$

- $\sum x^2 = 1239.56$
- $(\sum x)^2 = 165.4^2$
- $n = 23$

Standard Deviation; $s = \sqrt{\frac{n \sum x^2 - (\sum x)^2}{n(n-1)}} \approx \sqrt{\frac{23 \sum x^2 - (\sum x)^2}{23(23-1)}} \approx 1.51$

According to: $\bar{x} - t_{\alpha/2, n-1} s / \sqrt{n} \leq \mu \leq \bar{x} + t_{\alpha/2, n-1} s / \sqrt{n}$

- Class A:

$$n = 28 \Rightarrow n - 1 = 27 \Rightarrow t_{0.025, 27} = 2.052$$

$$s = 1.23 \rightarrow t_{0.025, 27} * s / \sqrt{n} = 2.052 * (1.23) / \sqrt{28} = 0.477$$

$$\Rightarrow 8.2 - 0.477 \leq \mu \leq 8.2 + 0.477$$

$$\Rightarrow 7.723 \leq \mu \leq 8.677$$

- Class B:

$$n = 23 \rightarrow n - 1 = 22 \rightarrow t_{0.025, 22} = 2.074$$

$$s = 1.51 \rightarrow t_{0.025, 22} * s / \sqrt{n} = 2.074 * (1.51) / \sqrt{23} = 0.653$$

$$\Rightarrow 7.191 - 0,653 \leq \mu \leq 7.191 + 0,653$$

$$\Rightarrow 6,538 \leq \mu \leq 7.844$$

Section 5.

-Giả thuyết được kiểm định điểm thi trung bình của sinh viên lớp A khác 5 và lớn hơn 5

$$\bar{x} = 7.97$$

$$s = 1.49$$

$$n = 30$$

a.

$$H_0: \mu = 5$$

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

$$t_0 = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{7.97 - 5}{1.49/\sqrt{30}} = 10.91$$

$$\Rightarrow \text{Bác bỏ } H_0 \text{ vì } t_0 > t_{\alpha/2; n-1} (10.91 > 2,05)$$

b.

$$H_0: \mu = 5$$

$$H_1: \mu > \mu_0$$

$$t_0 = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{7.97 - 5}{1.49/\sqrt{30}} = 10.91$$

$$\Rightarrow \text{Bác bỏ } H_0 \text{ vì } t_0 > t_{\alpha; n-1} (10.91 > 1,699)$$

Section 6.

-Giả thuyết được kiểm định kết luận điểm thi trung bình của sinh viên lớp B khác 5 và lớn hơn 5

$$\bar{x} = 6,12; s = 2.545; n = 30$$

a.

$$H_0: \mu = 5$$

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

$$t_0 = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{6,12 - 5}{2.545/\sqrt{30}} = 2.41$$

\Rightarrow Bác bỏ H_0 vì $t_0 > t_{\alpha/2; n-1}$ ($2.41 > 2,05$)

b.

$$H_0: \mu = 5$$

$$H_1: \mu > \mu_0$$

$$t_0 = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{6,12 - 5}{2.545/\sqrt{30}} = 2.41$$

\Rightarrow Bác bỏ H_0 vì $t_0 > t_{\alpha, n-1}$ ($2.41 > 1,699$)

Section 7.

-Kiểm định giả thiết tỷ lệ sinh viên nhóm A đạt điểm thi lớn hơn 5 (> 80%)

$$H_0: p = 0,8$$

$$H_1: p \neq 0,8$$

Test statistic:

$$\begin{aligned}
 z_0 &= \frac{x - np_0}{\sqrt{np_0(1-p_0)}} \\
 &= \frac{28 - 30 \cdot 0.8}{\sqrt{30 \cdot 0.8 \cdot (1 - 0.8)}} \\
 &\approx 1.83
 \end{aligned}$$

Reject H_0 : $p=0,8$ if $z_0 < -z_{0,025} = -1.96$

or $z_0 > z_{0,025} = 1.96$

=> kết luận: từ $z_0 \approx 1.83 \Rightarrow -z_{0,025} < z_0 < z_{0,025}$; chúng ta bác bỏ H_0 và kết luận tỷ lệ học sinh nhóm A có điểm kiểm tra trên 5 là khác 80% và cụ thể là lớn hơn 80%.

Section 8.

-Kiểm định giả thiết tỷ lệ sinh viên nhóm B đạt điểm thi lớn hơn 5 (> 80%)

$$H_0: p = 0,8$$

$$H_1: p \neq 0,8$$

Test statistic:

$$\begin{aligned}
 z_0 &= \frac{x - np_0}{\sqrt{np_0(1-p_0)}} \\
 &= \frac{23 - 30 \cdot 0,8}{\sqrt{30 \cdot 0,8(1 - 0,8)}} \\
 &\approx -0.456
 \end{aligned}$$

Reject H_0 : $p=0,8$ if $z_0 < -z_{0,025} = -1.96$

or $z_0 > z_{0,025} = 1.96$

=> kết luận: từ $z_0 \approx -0.456 \Rightarrow -z_{0,025} < z_0 < z_{0,025}$; chúng ta bác bỏ H_0 và kết luận tỷ lệ học sinh nhóm A có điểm kiểm tra

trên 5 là khác 80% và cụ thể là nhỏ hơn 80%.

Section 9.

-Với số liệu từ số liệu cho trong bảng ta thấy với bảng A có 30 học sinh và điểm trung bình của các em là 7,97. Đối với bảng B, chúng tôi có 30 học sinh và điểm trung bình của họ là 6,12.

Từ vấn đề, ta có:

$X1 = 7.97$; $X2 = 6,12$; $N1 = 30$; $N2 = 30$; $S1 = 1.49$; $S2 = 2.545$

$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}} = 47$$

If \bar{x}_1 , \bar{x}_2 , s_1^2 , and s_2^2 are the means and variances of two random samples of sizes n_1 and n_2 , respectively, from two independent normal populations with unknown and unequal variances, an approximate $100(1 - \alpha)\%$ confidence interval on the difference in means $\mu_1 - \mu_2$ is

$$\bar{x}_1 - \bar{x}_2 - t_{\alpha/2, v} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{x}_1 - \bar{x}_2 + t_{\alpha/2, v} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad (10-20)$$

where v is given by Equation 10-16 and $t_{\alpha/2, v}$ is the upper $\alpha/2$ percentage point of the t distribution with v degrees of freedom.

$$0.77 \leq \mu_1 - \mu_2 \leq 2.93$$

⇒ Kết luận: Điểm kiểm tra trung bình của nhóm A lớn hơn điểm của nhóm B

vì $X1 - X2 = 7,97 - 6,12 = 1,85$ vẫn nằm trong khoảng tin cậy 2 phía 95%.

6.	Difference in two proportions of two binominal parameters $p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$	$\hat{p}_1 - \hat{p}_2 - z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$ $\leq p_1 - p_2 \leq \hat{p}_1 - \hat{p}_2 + z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$
----	---	-------------------------	---

From the recipe $\Rightarrow -0,009 \leq P1 - P2 \leq 0,342$

Section 10.

Từ các giá trị ở trên ta đưa ra nhận xét:

1. Giá trị trung bình, độ lệch tiêu chuẩn, phân bố tần và biểu đồ cho dữ liệu điểm thi của hai nhóm

Nhóm A có giá trị trung bình = 7.97 ; độ lệch chuẩn của nhóm A (1.49)

Nhóm B có giá trị trung bình = 6,12 ; độ lệch chuẩn của nhóm B (2.545)

Giá trị trung bình của nhóm A (7.97) lớn hơn nhóm B (6,12) ; đồng thời độ lệch chuẩn của nhóm A (1.49) nhỏ hơn nhóm B với độ lệch chuẩn (2,545)

Từ đó, ta suy ra nhóm A có điểm số tốt hơn nhóm B và có sự biến động thấp hơn nhóm B. Tuy nhiên sự biến động không chênh lệch nhiều.

Nhóm A có điểm số nằm trong khoảng từ 4 đến 10, trong đó điểm từ 9~10 có tần xuất nhiều nhất (8) và điểm 7~8 cũng có tần xuất xuất hiện là 11, tần xuất dưới trung bình là 2.

Nhóm B có điểm số nằm trong khoảng từ 0 đến 10, trong đó điểm từ 9~10 có tần xuất nhiều nhất(3) và điểm 7~8 cũng có tần xuất xuất hiện là 2, tần xuất dưới trung bình là 7.

Nhóm A có điểm giỏi nhiều điểm giỏi hơn nhóm B.

Từ đó ta thấy rằng nhóm A có khả năng học tập tốt hơn nhóm B.

2. Khoảng tin cậy hai phía 95% cho điểm thi trung bình và điểm thi trung bình có điểm thi > 5 . (của hai lớp)

Chúng ta chắc chắn 95% (độ tin cậy) rằng trong khoảng $(7,7233 ; 8,347)$ có chứa giá trị trung bình của nhóm A

Chúng ta chắc chắn 95% (độ tin cậy) rằng trong khoảng $(5,167; 7,073)$ có chứa giá trị trung bình của nhóm B

Chúng ta chắc chắn 95% (độ tin cậy) rằng trong khoảng $(7.723 ; 8,677)$ có chứa điểm thi trung bình lớn hơn 5 của nhóm A

Chúng ta chắc chắn 95% (độ tin cậy) rằng trong khoảng $(6,538 ; 7.844)$ có chứa điểm thi trung bình lớn hơn 5 của nhóm B

3. Các giả thuyết được kiểm định

**Với giả thiết điểm thi trung bình của sinh viên lớp A:*

a. Bằng 5

Kết luận: Vì $(10,91 > 2,05)$ nên bác bỏ H_0 và kết luận điểm thi trung bình của sinh viên lớp A khác 5

b. Lớn hơn 5

Kết luận: Vì $(10,91 > 1,699)$ nên bác bỏ H_0 và kết luận điểm thi trung bình của sinh viên lớp A lớn hơn 5

*Với giả thiết điểm thi trung bình của sinh viên lớp B:

a. Bằng 5

Kết luận: Vì $(2.41 > 2.05)$ nên bác bỏ H_0 và kết luận điểm thi trung bình của sinh viên lớp B khác 5

b. Lớn hơn 5

Kết luận: Vì $(2.41 > 1.699)$ nên bác bỏ H_0 và kết luận điểm thi trung bình của sinh viên lớp B lớn hơn 5

*Kiểm định giả thiết tỷ lệ sinh viên nhóm A đạt điểm thi lớn hơn 5 (> 5)
= 50%

kết luận: từ $z_0 \approx 1.83 \Rightarrow -Z_{0,025} < Z_0 < Z_{0,025}$; chúng ta bác bỏ H_0 và kết luận tỷ lệ học sinh nhóm A có điểm kiểm tra

trên 5 là khác 80% và cụ thể là lớn hơn 80%.

*Kiểm định giả thiết tỷ lệ sinh viên nhóm A đạt điểm thi lớn hơn 5 (> 5)
= 30%

\Rightarrow kết luận: từ $z_0 \approx -0.456 \Rightarrow -Z_{0,025} < Z_0 < Z_{0,025}$; chúng ta bác bỏ H_0 và kết luận tỷ lệ học sinh nhóm A có điểm kiểm tra

trên 5 là khác 80% và cụ thể là nhỏ hơn 80%.

-Khoảng tin cậy 95% trên có ý nghĩa: hiệu số điểm của trung bình của 2 nhóm A và B nằm trong khoảng từ 0,77 đến 2,93

-Khoảng tin cậy 95% trên có ý nghĩa: tỷ lệ sinh viên đạt yêu cầu (điểm thi > 5) của 2 nhóm A và B nằm trong khoảng từ -0,009 đến 0,342

Tổng Quan: Về mọi mặt ta thấy chỉ số của nhóm A đều tốt hơn nhóm B và ta thấy được nhóm A có khả năng học tập tốt hơn nhóm B.