

# Project MAS291 – Group 2

## 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	4.4
4	6.6	19	9.7
5	5.8	20	10
6	4.6	21	4.1
7	6	22	4.3
8	5	23	8.5
9	4.4	24	6.5
10	4.2	25	9.5
11	9.8	26	7.5
12	9.1	27	6

<b>13</b>	<b>2</b>	<b>28</b>	<b>10</b>
<b>14</b>	<b>2.4</b>	<b>29</b>	<b>7.5</b>
<b>15</b>	<b>2.5</b>	<b>30</b>	<b>4.5</b>

## Class B

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

## III. Analysis

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

**- Group A mean and standard deviation:**

- $\sum x^2 = 1482,26$
- $(\sum x)^2 = 197,6^2$
- $n = 30$

$$\bar{x} \approx 6,567$$

$$S = \sqrt{\frac{n \sum x^2 - (\sum x)^2}{n(n-1)}} \approx \sqrt{\frac{30 \sum x^2 - (\sum x)^2}{30(30-1)}} \approx 2.496$$

**- Group B mean and standard deviation:**

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

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

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

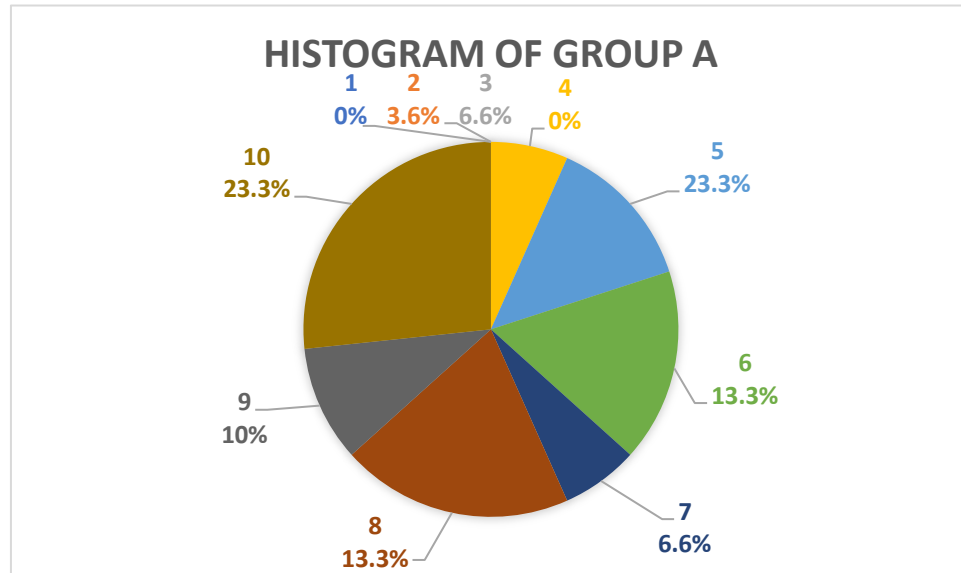
**Section 2. Build frequency distribution and histogram for test data of 2 groups**

**2.1. Frequency distribution (Phân bố tần số)**

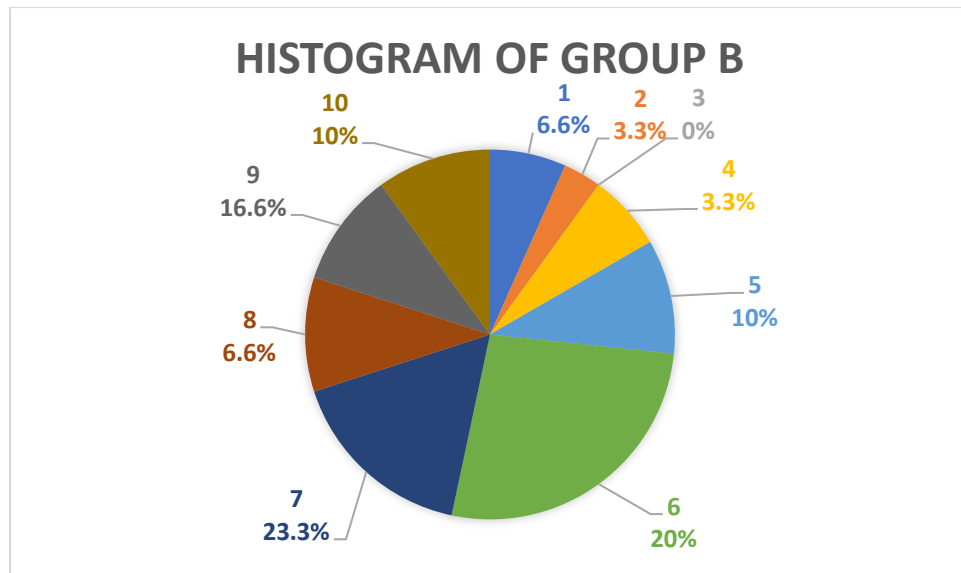
Mark	0 ~ 1	1 ~ 2	2 ~ 3	3 ~ 4	4 ~ 5	5 ~ 6	6 ~ 7	7 ~ 8	8 ~ 9	9 ~ 10
Frequency of Group A	0	1	2	0	8	3	2	4	3	7
Frequency of Group B	2	1	0	1	3	6	7	2	5	3

## 2.2. Histogram (Biểu đồ)

- A class chart:



- B class chart:



**Section 3.** Construct a 95% 2-sided confidence interval for the mean test scores of the 2 groups

**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:

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

$$s = 2.496 \Rightarrow t_{0,025;29} * s/\sqrt{n} = 2,05 * 2.496/\sqrt{30} = 0,934$$

$$\Rightarrow 6,567 - 0,934 \leq \mu \leq 6,567 + 0,934$$

$$\Rightarrow 5,633 \leq \mu \leq 7.501$$

- 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.**

If we're only interested in score of student which is above 5 then:

- Class A:

Sample mean:  $\bar{x} = 8.17$

Sample size:  $n = 19$

- $\sum x^2 = 1308.18$
- $(\sum x)^2 = 155.2^2$
- $n = 19$

$$\text{Standard Deviation; } s = \sqrt{\frac{n \sum x^2 - (\sum x)^2}{n(n-1)}} \approx \sqrt{\frac{19 \cdot 1308.18 - 155.2^2}{19(19-1)}} \approx 1.5$$

- Class B:

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

Sample size:  $n = 23$

- $\sum x^2 = 1239.56$
- $(\sum x)^2 = 165.4^2$
- $n = 19$

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

**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:

$$n = 19 \Rightarrow n - 1 = 18 \Rightarrow t_{0.025, 18} = 2.101$$

$$s = 1.5 \rightarrow t_{0.025, 18} * s / \sqrt{n} = 2.101 * (1.5) / \sqrt{19} = 0.723$$

$$\Rightarrow 8.17 - 0.723 \leq \mu \leq 8.17 + 0.723$$

$$\Rightarrow 7.447 \leq \mu \leq 8.893$$

- 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.

### -Hypothesis Tests on the Mean of Class A

$$\bar{x} = 6,567$$

$$s = 2.496$$

$$n = 30$$

**a.**

$$H_0: \mu = 5$$

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

$$t_0 = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{6,567 - 5}{2.496/\sqrt{30}} = 3.438$$

$\Rightarrow$  Reject because  $t_0 > t_{\alpha/2; n-1}$  ( $3.438 > 2,05$ )

**b.**

$$H_0: \mu = 5$$

$$H_1: \mu > \mu_0$$

$$t_0 = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{6,567 - 5}{2.496/\sqrt{30}} = 3.438$$

$\Rightarrow$  Reject because  $t_0 > t_{\alpha; n-1}$  ( $3.438 > 1,699$ )

## Section 6.

### -Hypothesis Tests on the Mean of Class B

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

**a.**

$$H_0: \mu = 5$$

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

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

$\Rightarrow$  Reject because  $t_0 > t_{\alpha/2; n-1}$  ( $2.41 > 2,05$ )

**b.**

$$H_0: \mu = 5$$

$$H_1: \mu > \mu_0$$

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

$\Rightarrow$  Reject because  $t_0 > t_{\alpha, n-1}$  ( $2.41 > 1,699$ )

## Section 7.

**State null and alternate hypothesis of class A:**

$$H_0: p = 0,5$$

$$H_1: p \neq 0,5$$

**Test statistic:**

$$\begin{aligned} z_0 &= \frac{x - np_0}{\sqrt{np_0(1-p_0)}} \\ &= \frac{19 - 30 \cdot 0,5}{\sqrt{30 \cdot 0,5 \cdot (1-0,5)}} \\ &\approx 1,46 \end{aligned}$$

**Reject  $H_0$ :**  $p=0,5$  if  $z_0 < -z_{0,025} = -1.96$

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

$\Rightarrow$  Conclusion: Since  $z_0 \approx 1,46 < z_{0,005} = 1.96$ ; we agree  $H_0$  and conclude the proportion of students in group A with test scores than 5 is more than 50%.



## Section 8.

State null and alternate hypothesis of class A:

$$H_0: p = 0,3$$

$$H_1: p \neq 0,3$$

Test statistic:

$$\begin{aligned} z_0 &= \frac{x - np_0}{\sqrt{np_0(1-p_0)}} \\ &= \frac{19 - 30 \cdot 0,3}{\sqrt{30 \cdot 0,3(1-0,3)}} \\ &\approx 3.98 \end{aligned}$$

Reject  $H_0$ :  $p=0,3$  if  $z_0 < -z_{0,025} = -1.96$

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

⇒ Conclusion: Since  $z_0 \approx 3.98 > z_{0,025} = 1.96$ ; we reject  $H_0$  and conclude the proportion of students in group A with test scores than 5 is not equal 30%.

## Section 9.

With the data from the data given in the table, we see that with table A there are 30 students and their average score is 6.567. As for Group B, we have 30 students and their average score is 6,12.

From the given problem, we have:

$$X1 = 6.567; X2 = 6,12; N1 = 30; N2 = 30; S1 = 2.496; S2 = 2.545$$

$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}} = 58 \text{ (If } v \text{ is not an integer, round down to the nearest integer)}$$

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.

$$-0,9 \leq \mu_1 - \mu_2 \leq 1,78$$

⇒ Conclusion: The average test score of group A is the same as that of group B because  $X1 - X2 = 6,567 - 6,12 = 0,447$ . It is still within the 99% 2-sided confidence interval.

6. Difference in two proportions of two binominal parameters  $p_1 - p_2$

$$\hat{p}_1 - \hat{p}_2$$

$$\begin{aligned} \hat{p}_1 - \hat{p}_2 - z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}} \\ \leq p_1 - p_2 \leq \hat{p}_1 - \hat{p}_2 + z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}} \end{aligned}$$

From the recipe ⇒  $-0,01469 \leq P1 - P2 \leq 0,2813$

## 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 = 6.567 ; độ lệch chuẩn của nhóm A (2.496)

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

Giá trị trung bình của nhóm A (6.567 ) lớn hơn nhóm B (6,12) ; đồng thời độ lệch chuẩn của nhóm A (2,496) 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 (9) và điểm 7~8 cũng có tần xuất xuất hiện là 8, tần xuất dưới trung bình là 6.

Nhóm B có điểm số nằm trong khoảng từ 0 đến 10, trong đó điểm từ 5~6 có tần xuất nhiều nhất(9) và điểm 6~7 cũng có tần xuất xuất hiện là 5, tần xuất dưới trung bình là 7.

Nhóm A có điểm giỏi(>8) nhiều điểm giỏi hơn nhóm B(13 > 9).

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 nhóm)**

Chúng ta chắc chắn 95% (độ tin cậy) rằng trong khoảng (5,633 ; 7.501) 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  $(7.447 ; 8.893)$  có chứa điểm thi trung bình lớn hơn 4 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  $(3.438 > 2,05)$  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  $(3.438 > 1,699)$  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: Vì  $z_0 \approx 1,46 < z_{0,005} = 1.96$  nên đồng ý  $H_0$  và kết luận tỉ lệ học sinh nhóm A có điểm thi trên 5 lớn hơn 50%.

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

Kết luận: Vì  $z_0 \approx 3.98 > z_{0,025} = 1.96$  nên bác bỏ  $H_0$  và kết luận tỉ lệ học sinh nhóm A có điểm thi trên 5 không bằng 30%.

### **Câu 9 :**

Khoảng tin cậy 99% trên có ý nghĩa: hiệu số điểm của trung bình của 2 nhóm A và B nằm trong khoảng từ -0,9 đến 1,78

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 -0,01469 đến 0,2813

**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.