FACULTY OF COMPUTING
UTM Johor Bahru

Semester II 2024/2025

Subject : SECI1143 PROBABILITY & STATISTICAL DATA ANALYSIS

Task : Chapter 5 (Hypothesis Testing for 2 Samples) & Chapter 6 (Chi-Square Test).

INSTRUCTION:

1. This is a **GROUP** assignment. Please clearly write the group members' names and matric numbers on the front page of the submission.
2. This assignment contributes to 5% of overall course marks.
3. Only **HANDWRITTEN** submission is accepted:
   a. Submissions using any reporting or statistical tools (e.g.: MS Word, MS Excel, etc.,) will be **REJECTED**.
   b. Make sure the submission is neatly written. Any submission with handwriting that is unreadable, will be **REJECTED**.
   c. For answers that need to draw graphs, using graph paper(s) is optional. You can use plain paper.
   d. Round your answers to **TWO** decimal places.
   e. Please scan/snapshot your work and save it as a PDF file (or follow any instruction from your lecturer).
4. Submission via eLearning – only **ONE** group member needs to submit on behalf of the group.

NAME	MATIRC NO.
IZWAN AZIZ BIN ISMAIL @ ABD MALEK	SX241894ECJHF01
YUARAJ A/L PARTHIPAN	SX241919ECRHF01
SITI NURNAJIHAH BINTI MOHAMAD ANUAR	SX232351ECRHF04
FATIN SYAHIRAH BINTI NOR RASHID	SX241920ECRHF01
ASWINI A/P CHANDRASAGARAN	SX242452ECRHF01

QUESTION 1 (10 marks)

A fitness coach wants to compare the effectiveness of two different training programs on weight loss.

- A group of 25 individuals following Program A lost an average of 4.3 kg after 8 weeks, with a standard deviation of 1.1 kg.

- Another group of 28 individuals following Program B lost an average of 5.1 kg, with a standard deviation of 1.3 kg.

Assuming the weight losses are approximately normally distributed and the population variances are not equal, test at the 0.05 significance level whether the mean weight loss for Program B is significantly greater than for Program A.

Question 1

Given data

- program A ⇒ $\bar{x}_A = 4.3$, $S_A = 1.1$, $n_A = 25$
- program B ⇒ $\bar{x}_B = 5.1$, $S_B = 1.3$, $n_B = 28$

1. Hypothesis statement :

$H_0 : \mu_B \le \mu_A$

$H_1 : \mu_B > \mu_A$

2. Given, $\alpha = 0.05$, the test statistic is :

$$t_0 = \frac{\bar{x}_B - \bar{x}_A - 0}{\sqrt{\frac{s_B^2}{n_B} + \frac{s_A^2}{n_A}}} \Rightarrow t_0 = \frac{5.1 - 4.3 - 0}{\sqrt{\frac{(1.3)^2}{28} + \frac{(1.1)^2}{25}}}$$

$$= \frac{0.8}{\sqrt{\frac{1.69}{28} + \frac{1.21}{25}}}$$

$$= \frac{0.8}{\sqrt{0.06 + 0.05}}$$

$$= \frac{0.8}{0.33}$$

$$= 2.42 \ \#$$

3. calculate the degree of freedom :

$$V = \frac{\left(\frac{S_B^2}{n_B} + \frac{S_A^2}{n_A}\right)^2}{\frac{\left(\frac{S_B^2}{n_B}\right)^2}{n_B - 1} + \frac{\left(\frac{S_A^2}{n_A}\right)^2}{n_A - 1}}$$

$$V = \frac{\left(\frac{(1.3)^2}{28} + \frac{(1.1)^2}{25}\right)^2}{\frac{\left(\frac{(1.3)^2}{28}\right)^2}{28-1} + \frac{\left(\frac{(1.1)^2}{25}\right)^2}{25-1}}$$

$$= \frac{(0.06 + 0.05)^2}{\frac{(0.06)^2}{27} + \frac{(0.05)^2}{24}}$$

$$= \frac{(0.11)^2}{\frac{0.0036}{27} + \frac{0.0025}{24}}$$

$$= \frac{0.0121}{0.000133 + 0.000104}$$

$$= 51.05$$

Therefore, using $\alpha = 0.05$ with $v = 51$

$$t_{0.05,51} = 1.676$$

Since $t_0 = 2.42 > 1.676$, we reject the null hypothesis. There is evidence to conclude that program B results in significantly greater weight loss than program A.

QUESTION 2 (8 marks)

A manufacturing engineer is testing the consistency of two machines used to fill bottles with soda. To determine whether the variability (standard deviation) of fill amounts is the same, she collects the following data:

- Machine A: A sample of 21 bottles has a standard deviation of 2.1 ml.
- Machine B: A sample of 16 bottles has a standard deviation of 1.5 ml.

At a 0.05 significance level, test whether there is a difference in the variances of the two machines' outputs. Assume the fill amounts are normally distributed.

Ref: Date:

Question 2

Machine A

$n_1 : 21$

$\sigma_1 : 2.1ml$

(variance) $\sigma_1^2 : 4.41 \, ml^2 \; (2.1^2)$

Machine B

$n_2 : 16$

$\sigma_2 : 1.5 \, ml$

(variance) $\sigma_2^2 : 2.25 \, ml^2 \; (1.5^2)$

① Hypothesis

Null hypothesis (H_0) : $\sigma_1^2 = \sigma_2^2$ (equal variances)

Alternative hypothesis (H_1) : $\sigma_1^2 \neq \sigma_2^2$ (unequal variances)

② Test Statistics

$$F = \sigma_1^2 / \sigma_2^2 \quad (\sigma_1^2 \text{ is the numerator to ensure } F \geq 1)$$

$$F = 4.41 / 2.25 = 1.96$$

③ Determine degrees of freedom

Machine A $df_1 : n_1 - 1 = 20$

Machine B $df_2 : n_2 - 1 = 15$

④ Critical values

i) Search for critical value of significance level /2 (0.025) with df_1 (20) & df_2 (15) in F distribution table.

$$F_{0.025,20,15} = 2.76 \text{ (upper critical value)}$$

ii) search for critical value of 1 - (significance level /2) (0.975) with df_1 (20) & df_2 (15) in F distribution table.

$$F_{0.975,20,15} = 1/(F_{0.025,15,20}) = 1/2.5731 = 0.39 \text{ (lower critical value)}$$

Ref: Date:

⑤ Conclusion

F value : 1.96

Upper critical value : 2.76

Lower critical value : 0.39

Since $0.39 < 1.96 < 2.76$, we fail to reject the null hypothesis, hence we have insufficient evidence to conclude that the variance of the two machine's outputs are different.

QUESTION 3 (15 marks)

A nutritionist wants to evaluate whether a new diet plan has a significant effect on cholesterol levels. She records the cholesterol levels (in mg/dL) of 10 patients before and after following the diet plan for 8 weeks. The data are as follows:

Patient	Before	After
1	220	200
2	210	190
3	215	212
4	225	210
5	230	215
6	218	205
7	212	200
8	208	198
9	220	210
10	214	205

a) Calculate the D, D^2 and their sums as in the table below. [7 marks]

Patient	Before	After	D	D^2
1	220	200	20	400
2	210	190	20	400
3	215	212	3	9

Patient	Before	After	D	D^2
4	225	210	15	225
5	230	215	15	225
6	218	205	13	169
7	212	200	12	144
8	208	198	10	100
9	220	210	10	100
10	214	205	9	81
			$\sum D = 127$	$\sum D^2 = 2{,}253$

b) Calculate the D and S_D. [2 marks]

c) At a 0.05 significance level, test whether the diet plan has a significant effect on cholesterol levels. [6 marks]

ANSWER :

Notes

Question 3

a)

Patient	Before	After	D(Before - After)	D^2
1	220	200	20	400
2	210	190	20	400
3	215	212	3	9
4	225	210	15	225
5	230	215	15	225
6	218	205	13	169
7	212	200	12	144
8	208	198	10	100
9	220	210	10	100
10	214	205	9	81
			$\Sigma D = 127$	$\Sigma D^2 =$ 2253

b) Calculate the [$\bar{D}$] and [s_D].

$$[\bar{D} = \frac{(\Sigma D)}{n} = \frac{127}{10} = 12.7]$$

$$[s\_D = \sqrt{\{[\Sigma D^2 - \frac{(\Sigma D)^2}{n}] / (n-1)\}}]$$

$$[s\_D = \sqrt{\{[2253 - \frac{(127)^2}{10} / (10-1)\}}]$$

$$[s\_D = \sqrt{\{\frac{[2253 - 1612.9]}{9}\}}]$$

$$[s\_D = \sqrt{\{\frac{640.1}{9}\}}]$$

$$[s\_D = \sqrt{71.1222}$$

$$[s\_D \approx 8.433$$

c) Null Hypothesis [$H\_0 : \mu\_D = 0$]

Alternative Hypothesis [$H\_1 : \mu\_D > 0$]

Test statistic: $(t = (\bar{D} - 0) / (s\_D / \sqrt{n})]$

$$[t = 12.7 / (\frac{8.433}{\sqrt{10}})]$$

$$[t = 12.7 / (\frac{8.433}{3.162})]$$

$$[t = \frac{12.7}{2.667}]$$

$$[t \approx 4.762]$$

Degrees of freedom $= [n-1 = 10-1 = 9]$

~~Using a t table calculator~~

Since the calculated t-value (4.762) is greater than the critical t-value (1.833), we reject the null hypothesis.

= The diet plan has significant effect on cholestrol levels at a 0.05 significance level.

QUESTION 4

A wildlife researcher wants to test whether the observed distribution of bird species in a forest follows a known regional distribution. Based on prior regional data, the expected proportions of four bird species are as follows:

- Sparrows: 40%
- Finches: 30%
- Robins: 20%
- Woodpeckers: 10%

She observes the following counts of birds during a field study:

Species	Observed Count
Sparrows	42
Finches	33
Robins	15
Woodpeckers	10

a) Calculate the difference between the observed (o) and expected count (ei) as in the table below:

Species	Observed Count	ei	$\frac{(o - e)^2}{e}$
Sparrows	42		
Finches	33		
Robins	15		
Woodpeckers	10		

[4marks]

b) Calculate the c2 test. [2 marks]

c) At the 0.05 significance level, test whether the observed distribution matches the expected proportions. [4 marks]

a) calculate different between observed (o) and expected (e) and compute $\frac{(o-e)^2}{e}$

① total all birds

$42 + 33 + 15 + 10 = 100$

②

species	OBSERVED (O)	EXPECTED (%)	(e) EXPECTED COUNT
① sparrows	42	40%	$0.40 \times 100 = 40$
② finches	33	30%	$0.30 \times 100 = 30$
③ robins	15	20%	$0.20 \times 100 = 20$
④ woodpeckers	10	10%	$0.10 \times 100 = 10$

③

species	observed (o)	expected (e)	$(o-e)^2/e$
① sparrows	42	40	$(42-40)^2/40 = 4/10 = 0.10$
② finches	33	30	$(33-30)^2/30 = 9/30 = 0.30$
③ robins	15	20	$(15-20)^2/20 = 25/20 = 1.25$
④ woodpeckers	10	10	$(10-10)^2/10 = 0/10 = 0.00$

b) calculate the chi-square (x^2) test statistics.

$$x^2 = \sum \frac{(o-e)^2}{e} = 0.10 + 0.30 + 1.25 + 0.00 = 1.65$$

c) At the 0.05 significance level, test whether the observed distribution matches the expected proportions.

① degrees of freedom

df = number of categories $- 1 = 4 - 1 = 3$

② critical value from chi-square table at $a = 0.05$, df $= 3$

$$x^2_{0.05,3} = 7.815$$

③ compare test statistic to critical value

$1.65 < 7.815$

QUESTION 5

A researcher wants to determine whether a preferred mode of transportation is associated with a person's type of employment. A random sample of 120 individuals is surveyed, and the results are recorded in the following table:

	Public Transport	Private Car	Bicycle	Total
Office Workers	20	30	10	60
Freelancers	15	20	5	40
Students	10	5	5	20
Total	45	55	20	120

At the 0.05 significance level, test whether type of employment and preferred mode of transportation are independent.

Aswini.

Assignment 3.

Question 5

$E_{ij} = \frac{(\text{Row Total}_i \times \text{Column Total}_j)}{\text{Grand Total}}$

$\chi^2 = \sum \frac{(O - E^2)}{E}$ → chi square Test (chapter 6 page 7).

		E_{ij}	χ^2
1	Office workers & Public Transport	$\frac{60 \times 45}{120} = 22.5$	$\frac{(20 - 22.5)^2}{22.5} = 0.28$
2	Office workers & Private Car	$\frac{60 \times 55}{120} = 27.5$	$\frac{(30 - 27.5)^2}{27.5} = 0.23$
3	Office workers & Bicycle	$\frac{60 \times 20}{120} = 10$	$\frac{(10 - 10)^2}{10} = 0$
4	Freelancers & Public Transport	$\frac{40 \times 45}{120} = 15$	$\frac{(15 - 15)^2}{15} = 0$
5	Freelancers & Private Car	$\frac{40 \times 55}{120} = 18.33$	$\frac{(20 - 18.33)^2}{18.33} = 0.15$
6	Freelancers & Bicycle.	$\frac{40 \times 20}{120} = 6.67$	$\frac{(5 - 6.67)^2}{6.67} = 0.42$
7	Students & Public Transport	$\frac{20 \times 45}{120} = 7.5$	$\frac{(10 - 7.5)^2}{7.5} = 0.83$
8	Students & Private Car	$\frac{20 \times 55}{120} = 9.17$	$\frac{(5 - 9.17)^2}{9.17} = 1.9$
9	Students & Bicycle.	$\frac{20 \times 20}{120} = 3.33$	$\frac{(5 - 3.33)^2}{3.33} = 0.84$

$\chi^2 = 0.28 + 0.23 + 0 + 0 + 0.15 + 0.42 + 0.83 + 1.9 + 0.84.$

$= 4.65$

$df = (3-1) \times (3-1) = 2 \times 2 = 4.$

the significance level is given as $= 0.05$.

$$\chi^2_{0.05,4} = 9.488$$

$$4.65 < 9.488$$

chi-square value is less than the critical value, the null hypothesis is not rejected.