ASSIGNMENT
BUM2413 APPLIED STATISTIC
SEMESTER I 2024/2025

SECTION: 13P   **GROUP NAME**: Standard Deviation

GROUP MEMBERS	STUDENT ID
1.MUHAMAD ALIFF AIMAN SHAHNI	CA21036
2.NUR LIYANA BINTI AZMAN	CA22030
3.NUR AMIRAH SHAHIRA BINTI ZULKIFLI	CA22044
4.NUR LIYANA BINTI HAMZAH	CA22069
LECTURER : LAILA AMERA BINTI AZIZ	
SUBMISSION DATE : 7/1/2025	

FOR EXAMINER USE ONLY					
Question	**Marks**	**Your Marks**	**Question**	**Marks**	**Your Marks**
1	**2**		**8**	**2**	
2	**1**		**9**	**2**	
3	**2**		**10**	**4**	
4	**7**		**11**	**24**	
5	**3**		**12**	**1**	
6	**4**				
7	**8**				
			TOTAL	**60**	

1. **Identify** a topic/ problem that you are interested in studying. Provide a brief **description** of your study and state at least **ONE (1)** study objective.

   Due to the increasing popularity of digital streaming platforms and online gaming, Malaysians are subscribing to multiple online entertainment services. However, the rise in subscription-based entertainment raises questions about affordability and spending habits among individuals and households. Therefore, this study aims to collect data on the monthly spending of Malaysians on online entertainment subscriptions to understand their financial commitment and assess trends in digital consumption. This information can also shed light on consumer priorities and their potential financial impact.

   (2 Marks)

2. State the **population** of the study

   Malaysians

   (1 Mark)

3. Determine a **single quantitative variable** that is related to your chosen problem. Identify the **type of level of measurement** for the variable.

   Quantitative variable : Monthly expenditure on online entertainment subscriptions (in Malaysian Ringgit, MYR).
   Level of measurement : Ratio level
   The variable is measured on a ratio scale because it has a true zero point (zero expenditure indicates no subscription spending) and allows for meaningful comparisons
   Example, someone spending MYR 100 spends twice as much as someone spending MYR 50.

   (2 Marks)

4. Divide the data collected into **two significant groups** (e.g.: gender (male/female), faculty, year of study, etc.) that are related to the study. The sample size is **at least 50 observations** for **each group**.

(i) State the **name of the groups**.

Malaysian	
Group 1	**Group 2**
Students	Non-Students

(1 Mark)

(ii) Present the data collected according to the groups in a table.

Malaysian	Students	
	Data	Value (Money RM)
	1	80.0
	2	200.0
	3	300.0
	4	60.0
	5	0.0
	6	100.0
	7	50.0
	8	0.0

	9	30.0
	10	100.0
	11	0.0
	12	20.0
	13	45.0
	14	11.0
	15	115.0
	16	8.9
	17	50.0
	18	500.0
	19	0.0
	20	50.0
	21	0.0
	22	50.0
	23	35.0
	24	100.0
	25	0.0
	26	70.0
	27	0.0
	28	60.0

	29	10.0
	30	0.0
	31	40.0
	32	3.0
	33	25.0
	34	200.0
	35	17.10
	36	250.0
	37	0.0
	38	50.0
	39	0.0
	40	0.0
	41	5.65
	42	80.0
	43	100.0
	44	100.0
	45	40.0
	46	50.0
	47	0.0
	48	30.0

	49	30.0
	50	35.6

Malaysian	Non-Students	
	Data	Value (Money RM)
	1	500.0
	2	150.0
	3	0.0
	4	0.0
	5	150.0
	6	50.0
	7	100.0
	8	100.0
	9	7.0
	10	400.0
	11	50.0
	12	100.0
	13	0.0
	14	350.0

	15	150.0
	16	0.0
	17	0.0
	18	22.0
	19	50.0
	20	100.0
	21	70.0
	22	50.0
	23	80.0
	24	50.0
	25	0.0
	26	35.0
	27	100.0
	28	100.0
	29	1.0
	30	50.0
	31	30.0
	32	80.0
	33	27.1
	34	48.0

	35	100.0
	36	50.0
	37	50.0
	38	100.0
	39	15.8
	40	50.0
	41	100.0
	42	0.0
	43	100.0
	44	62.9
	45	55.0
	46	150.0
	47	60.0
	48	120.0
	49	50.0
	50	40.0
	51	70.0
Total	50 (Students) + 51 (Non-Students) = 101	

(2 Marks)

(iii) Identify the **method of data collection** being used. Provide significant **evidence**.

Methods of data collection is by using Questionnaire by Google form.

Online Entertainment

How much do you spent on online entertainment per month? (in RM) *

Example : RM50

Your answer

What kind of entertainment do you spent? *

☐ None
☐ Spotify
☐ Netflix
☐ Apple Music
☐ Viu
☐ Youtube Premium
☐ Prime Video
☐ Disney+ Hotstar
☐ Others

Back | Submit | Clear form

(2 Marks)

(iv) State the **sampling method** you use to collect the data. **Explain** the sampling method process.

Sampling Methods: Volunteering Sampling

We developed a Google Form comprising several questions for Malaysian citizens regarding their status, preferred types of internet entertainment, and monthly expenditures on online entertainment. The Google Form is disseminated to the public by sharing it through WhatsApp and Telegram groups.

(2 Marks)

5. For each set of data, obtain the **descriptive statistics** using *Microsoft Excel*. Then, summarise the **measures of central tendency (mean, median, mode, midrange)** and **measures of variation (range, standard deviation, variance)** in the following table.

Student		Non-Student	
Mean	62.025	Mean	81.83921569
Standard Error	12.86263171	Standard Error	13.44630486
Median	37.8	Median	50
Mode	0	Mode	50
Standard Deviation	90.95254103	Standard Deviation	96.02582377
Sample Variance	8272.364719	Sample Variance	9220.958831
Kurtosis	11.04268508	Kurtosis	9.154615813
Skewness	2.988180145	Skewness	2.816967214
Range	500	Range	500
Minimum	0	Minimum	0
Maximum	500	Maximum	500
Sum	3101.25	Sum	4173.8
Count	50	Count	51

Measure Of Central Tendency		
	Student	**Non-Student**
Mean	62.025	81.8392
Median	37.8	50
Midrange	250	250
Mode	0	50
Measure of Variation		
	Student	**Non-Student**

| Standard Deviation	90.9525	96.0258
Sample Variance	8272.3647	9220.9588
Range	500	500

(3 Marks)

6. **Compare** and **comment** on the measures of central tendency and measures of variation between **Group 1** and **Group 2**.

The measures of central tendency for Non-Students show a higher mean (RM81.8392), median (RM50), and mode (RM50) compared to Students (mean RM62.025, median RM37.8, mode RM0). In terms of variation, Non-Students exhibit a slightly higher standard deviation (96.0258) and sample variance (9220.9588) than Students (standard deviation 90.9525, sample variance 8272.3647). Both groups have an identical range (500). This suggests that Non-Students tend to have higher and more consistent values, while Students display slightly less variability and lower central tendencies, with spending concentrated around smaller values.

(4 Marks)

7. Construct **box plots** for the **two** sets of data on the **same axis**. Identify the **shape of the distribution** for each boxplot. **Compare** and **comment** on the **average** and **variability** of the boxplots.

Boxplot
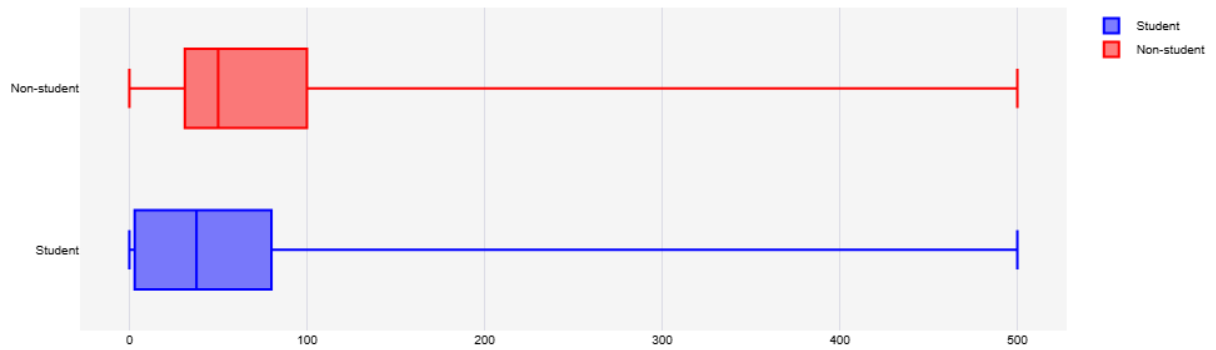
Data Set	Shape of Distribution
Non-student	Right-skewed distribution
Student	Right-skewed distribution

Compare and comment on the average:

Non-students have a higher average (mean: 81.8392) compared to students (mean: 62.025), indicating that non-students tend to have larger values overall. In terms of variability, students show slightly greater spread in the middle 50% of the data (IQR: 77 vs. 68.75 for non-students). Both groups have right skewed distributions, with students exhibiting a slightly stronger skewness (2.9882 vs. 2.817) and higher kurtosis (11.0427 vs. 9.1546), indicating more extreme values and asymmetry. Overall, students' data appears more variable and has more pronounced outliers compared to non-students.

(8 Marks)

8. What is the **best measure of central tendency** to describe your data? Give a **reason**.

The best measure of central tendency to describe this data is the **median** because the median is less affected by extreme values and outliers, which are prominent in both datasets (e.g., outliers of 200, 300, and 500 for students; 350, 400, and 500 for non-students). Unlike the mean, which can be heavily influenced by these outliers, the median provides a better representation of the central location for skewed data.
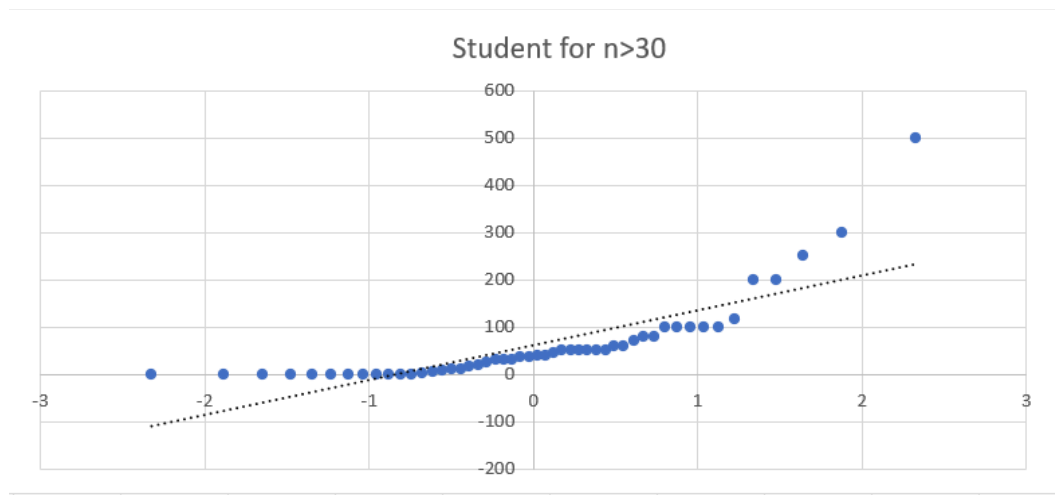
(2 Marks)

9. What is the **best measure of variation** to describe your data? Give a **reason**.

The **interquartile range (IQR)** is the best measure of variation to describe this data because it focuses on the spread of the middle 50% of the data and is not influenced by extreme values or outliers, which are present in both datasets. Unlike the standard deviation, which can be distorted by the highly skewed nature of the data and the outliers, the IQR provides a more robust and reliable measure of variation for this context.

(2 Marks)

10. Construct a **normal probability plot** for each data set. Do the data appear to come from an approximately normal distribution?

Since the data did not lie approximately on a straight line, the data appear to come from not a normal distribution.
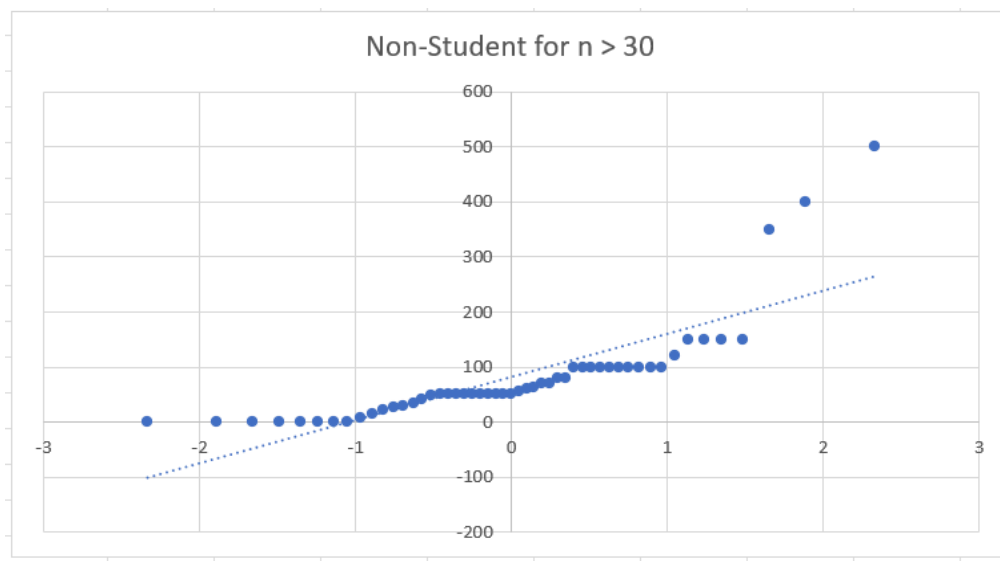
Since the data did not lie approximately on a straight line, the data appear to come from not a normal distribution.

(4 Marks)

In Chapter 3, you have learnt statistical hypothesis testing concerning a parameter(s) of one and two populations. Hypothesis testing is one of the inferential statistics in statistical analysis. The parameters are the population mean, proportion, variance and standard deviation. Assuming that the data obtained in (4) is a normally distributed population, answer the following questions using the **P-value approach and Microsoft Excel.**

(**NOTE**: Create your own hypothesised mean with justification, may use the overall mean of the data)

a. Create a situation and conduct **hypothesis testing for one population mean** from one of the groups.

A study was conducted to investigate whether the average monthly expenditure on online entertainment subscriptions by students is significantly different from RM50. A sample of 50 students was selected, and their expenditures were recorded. Test this claim at a 5% significance level.

Step 1

$H_0: \mu = 50$

$H_1: \mu \neq 50 (claim)$

Step 2

z-Test: One Sample for Means	
	STUDENT
Mean	42.84090909
Known Variance	25
Observations	11
Hypothesized Mean	50
z	-4.748699207
P(Z<=z) one-tail	1.02365E-06
z Critical one-tail	1.644853627
P(Z<=z) two-tail	2.04729E-06
z Critical two-tail	1.959963985

Step 3

P-value $= 2.0473 \times 10^{-6}$. The test is between right-tailed and left- tailed.

Step 4

Since (P-value $= 2.0473 \times 10^{-6} < \alpha = 0.05$)

Then, reject H_o

Step 5

At $\alpha = 0.05$, there is sufficient evidence to support the claim.

(8 Marks)

b. Choose one probability sampling method to select less than 30 data from each group.

i. Identify which sampling method you choose to select the data and explain the sampling method process.

Random sampling technique. List all data and generate 25 random numbers based on their group (Student and Non-Student).

(2 Marks)

ii. Present the selected data in a table.

Student	Non-Student
5.65	50
50	70
50	0
11	15.8
0	22
0	120
5.65	100
45	100
100	80
0	48
0	7
50	0
11	100
0	0

0	0
50	60
0	7
10	150
0	150
80	100
50	48
100	70
0	35
25	100
115	50

(2 Marks)

c. Create a situation to conduct a **hypothesis testing** using the data selected in (b) to **compare two population means** between the groups**.**

A study was conducted to investigate whether there is a significant difference in the average monthly expenditures on online entertainment subscriptions between students and non-students. A random sample of 25 students and 25 non-students was selected. Their monthly expenditures (in RM) are as recorded in question 11(b). Test this claim that the means of non-students spending online entertainment subscriptions per month is more than student at 5% significance level

Hypothesis for variance equality:

Step (i)

$$H_0 : \sigma_S^2 = \sigma_{NS}^2 \ (claim)$$ S : Student

$$H_1 : \sigma_S^2 \neq \sigma_{NS}^2$$ NS : Non-Student

Step (ii)

Random Student		Random Non-Student			
5.65		50		p-value	0.230284929
50		70			
50		0			
11		15.8			
0		22			
0		120			
5.65		100			
45		100			
100		80			
0		48			
0		7			
50		0			
11		100			
0		0			
0		0			
50		60			
0		7			
10		150			
0		150			
80		100			
50		48			
100		70			
0		35			
25		100			
115		50			

Step (iii)

Since(P-value = 0.2303) > (α = 0.05), we do not reject H$_o$.

Step(iv)

At α = 0.05, there is insufficient evidence to support reject the claim. Therefore there is no difference in variability.

Hypothesis for two population means:

Step (i)

$$H_0 : \mu_{NS} - \mu_S \le 0$$

$$H_1 : \mu_{NS} - \mu_S > 0 \; (claim)$$

Step (ii)

t-Test: Two-Sample Assuming Equal Variances		
	STUDENT	NON-STUDENT
Mean	30.332	59.312
Variance	1340.837058	2204.866933
Observations	25	25
Pooled Variance	1772.851996	
Hypothesized Mean	0	
df	48	
t Stat	-2.433420416	
P(T<=t) one-tail	0.009365304	
t Critical one-tail	1.677224196	
P(T<=t) two-tail	0.018730607	
t Critical two-tail	2.010634758	

Step (iii)

P-value = 0.0094. The test is right- tailed.

Step (iv)

Since(P-value = 0.0094) < (α = 0.05), we reject H$_o$.

Step (v)

At α = 0.05, there is sufficient evidence to support the claim. Therefore, non-students spending online entertainment subscriptions per month is more than student

(12 Marks)

11. Based on your problem/topic stated in (1), give any relevant **conclusion** for the study.

Based on the issue mentioned in (1), we discovered that Malaysians spend quite a bit of their monthly money on subscriptions to online entertainment. This could put a strain on finances, particularly for households with lower incomes, and make it more difficult for them to prioritise other important needs.

(1 Mark)