TEST 1

NAME	
COURSE CODE	DUM 2413 STATISTICS AND PROBABILITY
DURATION	1 HOUR AND 30 MINUTES

QUESTION 1

Mr. Adam is the car sales manager, who interested to investigate the level of satisfaction among customers who purchased the cars for the past 12 months. From the sales of 480 customers, Mr. Adam has randomly selected 20 customers from each brand of refrigerators. **Table 1** shows the number of customers for each brand of cars.

Table 1

Brand of cars	**Numbers of customers**
A	150
B	90
C	120
D	40
E	80

(i) State the population of interest and the type of population involved in this study.

(ii) State the variable of interest and level of measurement of the variable involved in this study.

(iii) What is the sampling method used by Mr. Adam?

QUESTION 2

Before Halloween, Mr. Hafiz purchased two bags of chocolate bars, where there are 19 and 23 chocolate bars in the first and second bags, respectively. He is interested to study the distribution of the weight (n grams) of chocolate bars, therefore he has mixed out and weighed the two bags of chocolate yielded the following weights record as illustrated in **Table 2**.

Table 2

19.8	20.3	16.8	19.6	17.5	19.4	15.4
18.9	17.5	20.8	18.8	15.5	17.8	18.0
15.4	17.6	15.7	18.4	17.3	16.7	15.6
14.2	15.9	16.8	15.2	15.0	16.0	16.4
15.4	16.0	16.1	17.0	16.0	15.1	14.0
15.8	15.7	16.6	15.2	15.5	14.9	15.3

(i) Complete the table below.

Class Limits	**Class Boundaries**	**Midpoint**	**Tally**	**Frequency**
14.0-14.9				
15.0-15.9				
16.0-16.9				
17.0-17.9				
18.0-18.9				
19.0-19.9				
20.0-20.9				

(ii) Based on the information in **(i)**, find the mean, mode and median. Then, identify the distribution of the weights of chocolate bars.

(iii) Based on the information in **(i)**, construct a histogram of the weights of chocolate bars on the graph paper. Would the distribution shape of the chocolate bars weight determined in (ii) can be supported by the histogram? Justify your answer.

QUESTION 3

A manufacturing engineer interested to compare the efficiency of two different saws, X and Y manufacturer by his company, which used to cut columns for a gazebo. To pursue his objective, he measured the diameters (in centimeters) of 9 columns cut using the saws X and Y, respectively as recorded in **Table 3**.

Table 3

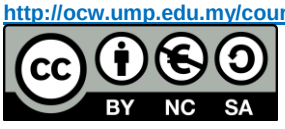
Saw X	8.00	8.02	8.02	8.04	8.04	8.04	8.07	8.10	8.11
Saw Y	8.04	8.04	8.06	8.06	8.07	8.08	8.08	8.10	8.10

(i) Find the sample means and standard deviations of each set of measurements. Then, compares the variation of the both measurements.

(ii) The following table shows important information for construct boxplots, including the five-number summaries and outlier(s) of the boxplots for each set of measurements. Complete the table below by showing all necessary calculations.

Information	**Saw** X	**Saw** Y
Minimum		
First Quartile		
Median		
Third Quartile		
Maximum		
Outlier		

(iii) Based on the answer in **(ii)**, construct boxplots for each set of measurements on same axis on the graph. Then, compares the **shape** and **variation** of the both measurements.



QUESTION 4

On the first day of Statistics & Probability lecture last semester, 50 students are randomly selected based on the first number of their student ID to answer that the time taken from their hostel to lecture hall (to the nearest minutes). The collected data is depicted in **Table 4**.

Table 4

Distance	**Midpoint**	**Frequency**
$1 \le x \le 4$	2.5	4
$5 \le x \le 8$	6.5	7
$9 \le x \le 12$	10.5	12
$13 \le x \le 16$	14.5	15
$17 \le x \le 20$	18.5	11
$21 \le x \le 24$	22.5	9
$25 \le x \le 28$	26.5	2

(i) Given that the mean of time taken by students from their hostel to lecturer hall is 14.3000 minutes, determine the standard deviation time taken based on the collected data depicted in **Table 4**.

(ii) According to Chebyshev's theorem, at least what percent of the time taken by students from their hostel to lecture hall is between 1.9393 and 26.6607 minutes?

END OF QUESTION PAPER

APPENDIX-TABLE OF FORMULAS

MEASURES OF CENTRAL TENDENCY

Ungrouped Data

$\mu = \dfrac{\sum_{i=1}^{N} x_i}{N}$; N is population size.

$\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{N}$; n sample size.

Grouped Data

$\bar{x} = \dfrac{\sum f_i x_i}{\sum f_i}$ where f_i is frequency and x_i is midpoint.

$\text{Median} = \text{L}_{\text{median}} + \left(\dfrac{\left(n/2\right) - f_{\text{L}}}{f_{\text{Median}}} \right) * \text{C}$

$\text{Mode} = \text{L}_{\text{mode}} + \left(\dfrac{\lambda_1}{\lambda_1 + \lambda_2} \right) * \text{C}$

MEASURES OF VARIATION

Ungrouped Data

$\sigma = \sqrt{\dfrac{\sum_{i=1}^{N} \left(x_i - \mu\right)^2}{N}}$

$s = \sqrt{\dfrac{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2}{n-1}}$

$\sigma^2 = \dfrac{\sum_{i=1}^{N} \left(x_i - \mu\right)^2}{N}$

$s^2 = \dfrac{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2}{n-1}$

Grouped Data

$\sigma = \sqrt{\dfrac{\sum_{i=1}^{k} f_i \left(x_i - \mu\right)^2}{\sum_{i=1}^{k} f_i}}$; k is number of classes

$s = \sqrt{\dfrac{\sum_{i=1}^{k} f_i \left(x_i - \bar{x}\right)^2}{\left(\sum_{i=1}^{k} f_i\right) - 1}}$

$\sigma^2 = \dfrac{\sum_{i=1}^{k} f_i \left(x_i - \mu\right)^2}{\sum_{i=1}^{k} f_i}$

$s^2 = \dfrac{\sum_{i=1}^{k} f_i \left(x_i - \bar{x}\right)^2}{\left(\sum_{i=1}^{k} f_i\right) - 1}$

Chebyshev's Theorem

At least $\left(1 - \dfrac{1}{k^2}\right)$ data fall within $\left(\mu \pm k\sigma\right)$ @ At least $\left(1 - \dfrac{1}{k^2}\right)$ data fall within $\left(\bar{x} \pm ks\right)$

$Q_i = \text{L} + \left(\dfrac{\left(in/4\right) - f_{\text{L}}}{f_{Q_i}} \right) * \text{C}$

$D_i = \text{L} + \left(\dfrac{\left(in/10\right) - f_{\text{L}}}{f_{D_i}} \right) * \text{C}$

$P_i = \text{L} + \left(\dfrac{\left(in/100\right) - f_{\text{L}}}{f_{P_i}} \right) * \text{C}$

OUTLIER

$x < Q_1 - 1.5\left(Q_3 - Q_1\right)$ and $x > Q_3 + 1.5\left(Q_3 - Q_1\right)$