DATA SCIENCE

2025

YEAR 1

KAPASA MAKASA UNIVERSITY

SCHOOL OF APPLIED SCIENCE AND OPEN LEARNING

DEPARTMENT OF INFORMATION AND COMMUNICATION TECHNOLOGY

FINAL EXAMINATION

Course Title: **Introduction to Statistics**

Course Code: **DS 110/ISB 140**

Date: **October, 2025**

Duration: **3 HOURS**

Total Marks: **60**

GENERAL INSTRUCTIONS TO CANDIDATES:

1. There are **Six** (6) Questions in this Examination paper. Attempt any **Four** (4) Questions.
2. Show all essential working to earn full marks.
3. Each Question carries **15 Marks** and the total mark for the **Four** (4) attempted Questions is 60.
4. You may use a non-programmed calculator in this Examination
5. Start each answer on a new page.
6. Write clearly and readably.

Please Note: *Students that are caught with foreign material, copying or behave in any manner that is viewed as examination malpractice/misconduct will be subjected to a disciplinary hearing.*

QUESTION ONE

(a) Define the following terms:

(i) A Parameter, [1]

(ii) A population, [1]

(iii) A variable. [1]

(b) A company records the response time (in seconds) of its web servers when processing online transactions. A sample of 125 response times was collected and grouped into the following intervals:

Response Time (sec)	Frequency
0.25 – 0.28	6
0.28 – 0.29	12
0.29 – 0.30	27
0.30 – 0.31	30
0.31 – 0.32	18
0.32 – 0.33	14
0.33 – 0.34	9
0.34 – 0.35	4
0.35 – 0.40	5

(i) Superimpose the frequency polygon on the histogram to represent the data above, [3]

(ii) Compute the mode response times, [4]

(iii) Compute the standard deviation of the response times. [5]

QUESTION TWO

(a) (i) Define the two types of errors in hypothesis testing. [2]

(ii) State the Central Limit Theorem. [2]

(iii) The lifetime of laptop batteries is normally distributed with a mean of 13 hours and a standard deviation of 1.5 hours. What is the probability that a sample of 25 batteries will have an average lifetime between 12.5 and 14 hours? [3]

(b) (i) For a skewed distribution, the mean is 16, the median is 20 and the standard deviation is 5. Calculate the Pearson's coefficient of skewness and sketch its curve. [3]

(ii) Given that $e = 2.3$, $\sigma = 15.40$, confidence level $= 99\%$. Determine the sample size for the estimate of μ. [2]

(iii) If $Z \sim N(0,1)$. Find the value of a, where $P(Z > a) = 0.235$. [3]

QUESTION THREE

Suppose three lecturers teach separate Data Science classes. The final scores of 15 students are recorded as follows:

Lecturer 1	Lecturer 2	Lecturer 3
73	89	82
43	80	88
78	48	91
51	85	68
77	56	71

(a) Prepare an ANOVA table for the data. [7]

(b)Test at $\alpha = 0.05$ whether the results of the three classes are significantly different. [3]

(c) Construct a 90% confidence interval for the mean difference between Lecturer 2 and lecturer 3 scores, given that $\bar{x}_2 = 73.5, n_2 = 6, s_2 = 19.10759, \bar{x}_3 = 68, n_3 = 4, \ s_3 = 8.83176$. [5]

QUESTION FOUR

(a)The following dataset represents customer satisfaction scores for an e-commerce platform:

58, 52, 68, 86, 72, 66, 97, 89, 84, 91, 91, 92, 66, 68, 87, 86,
73, 61, 70, 75, 72, 73, 85, 84, 90, 57, 77, 76, 84, 93, 58, 47

(i) Use a stem-and-leaf display to arrange the dataset. [2]

(ii) Construct a boxplot and comment on the distribution. [6]

(iii) Compute the quartile coefficient of skewness. [2]

(b) In a survey of 400 business executives, 90 said they are willing to share their company's annual data openly. Find:

(i) A point estimate of the proportion willing to share. [2]

(ii) A 95% confidence interval for this proportion. [3]

QUESTION FIVE

(a) The numbers 3·2, 5·8, 7·9 and 4·5 have frequencies x, $(x + 2)$, $(x - 3)$ and $(x + 6)$, respectively. If their arithmetic mean is 4·876, find

(i) the value of x. [2]

(ii) the Geometric Mean (GM), [2]

(iii) the Harmonic Mean (HM). [2]

(b) Two machine learning models are tested on different datasets selected from two normally distributed populations with unknown but equal variances. Their results are shown below:

Sample	Size (n)	Mean ($\bar{x}$)	Std Dev (S)
Model 1	16	2.4	0.1
Model 2	11	2.6	0.5

(i) Construct a 95% confidence interval for the variance ratio σ_1^2/σ_2^2. [3]

(ii) Are the population variances of the two models different? justify. [2]

(iii) Test at $\alpha = 0.1$ whether the variances are equal. [4]

QUESTION SIX

(a) A data analytics company studied how mobile app downloads (Y) depend on the age of users (X). A sample of 7 users is shown below:

Age (X)	8	3	6	9	2	5	60
Downloads (Y)	45	210	100	33	267	134	109

The summary of data is given below

$$\sum_{i=1}^{n} x = 93 \qquad \sum_{i=1}^{n} x^2 = 3819 \qquad n = 7$$

$$\sum_{i=1}^{n} y = 898 \qquad \sum_{i=1}^{n} y^2 = 158340 \qquad \sum_{i=1}^{n} xy = 9631$$

$$\bar{y} = 128.286 \qquad \bar{x} = 13.286$$

(i) Draw a scatter diagram and comment on the linear relationship between age and downloads. [3]

(ii) Estimate the regression line. [4]

(b) In a survey of 300 adults on digital privacy, responses were classified as follows:

	In Favor (F)	Against (A)	No Opinion (N)
Men (M)	93	70	12
Women (W)	87	32	6

(i) Calculate the expected frequencies, assuming independence between gender and opinion. [3]

(ii) Test at 1% significance whether gender and opinion are dependent. [5]

END OF EXAMINATION

STATISTICAL FORMULAS

UNGROUPED DATA

1 **Mean:** $\bar{x} = \frac{1}{n}\sum_{i=1}^{n}(x_i)$

2. **Standard Deviation:** $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$

$$= \sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{(\sum x_i)^2}{n}\right]} = \sqrt{S^2}$$

GROUPED DATA

1. **Mean:** $\bar{X} = \frac{\sum fx}{\sum f}$

2. **Mode:** $Mode = L_{m_o} + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right) i$

3. **Standard Deviation:** $S = \sqrt{\frac{1}{\sum f - 1}\left[\sum_{i=1}^{n} f x_i^2 - \frac{(\sum f x_i)^2}{\sum f}\right]} = \sqrt{S^2}$

LINEAR REGRESSION AND ANOVA

(1) $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$ (2) $S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n}$

(3) $S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}$ (4) $\hat{\beta} = \frac{S_{xy}}{S_{xx}}$ (5) $SST = \sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}^2 - \frac{y_{..}^2}{N}$

(6) $SSTrt = \sum_{i=1}^{k} \frac{y_{i.}^2}{n} - \frac{y_{..}^2}{N}$

CONFIDENCE INTERVAL ESTIMATION

One Sample : $C.I: \bar{x} \pm z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} = \left(\bar{x} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$ *for* $z -$ *distribution*

$$C.I: \left(\bar{x} - t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}}, \bar{x} + t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}}\right) \text{ for } t - \text{distribution}$$

$$C.I: \left(\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right), \text{for } \chi^2 - \text{distribution}$$

Two samples:

(1) Dependent Samples: $C.I:\ \bar{d} \pm t_{\frac{\alpha}{2},n-1}\frac{s_d}{\sqrt{n}}$

(2) Independent Samples:

(a) $C.I: (\bar{x}_1 - \bar{x}_2) \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$ for known σ_1^2 and σ_2^2, $n_1, n_2 \geq 30$ *or* $n_1, n_2 < 30$

(b) $C.I: (\bar{x}_1 - \bar{x}_2) \pm z_{\frac{\alpha}{2}}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$ *for unknown* σ_1^2 *and* σ_2^2, $n_1, n_2 \geq 30$

(c) $C.I: (\bar{x}_1 - \bar{x}_2) \pm t_{\frac{\alpha}{2},\nu}\sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$ *for unknown* σ_1^2 *and* $\sigma_2^2, n_1, n_2 < 30$

$$\nu = n_1 + n_2 - 2$$

(d) pooled variance: $s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$

Two Sample Proportion: (1) $C.I: (\hat{p}_1 - \hat{p}_2) \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$, $n_1 + n_2 \geq 30$

(2) $C.I: (\hat{p}_1 - \hat{p}_2) \pm t_{\frac{\alpha}{2}}(n_1 + n_2 - 2)\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$, $n_1 + n_2 < 30$

Two sample Ratio of Variances : $C.I:\ \frac{s_2^2}{s_1^2}\frac{1}{f_{\frac{\alpha}{2}}(\nu_2,\nu_1)} < \frac{\sigma_2^2}{\sigma_1^2} < \frac{s_2^2}{s_1^2}\frac{1}{f_{1-\frac{\alpha}{2}}(\nu_2,\nu_1)}$, and

$$f_\alpha(\nu_2,\nu_1) = \frac{1}{f_{1-\alpha}(\nu_1,\nu_2)}$$

HYPOTHESIS TESTING APPROACH: TEST STATISTIC

One Sample

(a) $Z = \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$ (b) $T = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}} \sim T_{n-1}$ (c) Proportion: $Z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$

(d) Variance: $\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$

Two Samples

(a) Independent: (1) $Z = \frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$ (2) $T = \frac{(\bar{x}_1-\bar{x}_2)-\mu_0}{\sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim t_{\nu=n_1+n_2-2}$

(b) Dependent: $T = \frac{\bar{D}-\mu_0}{\frac{s_d}{\sqrt{n}}} \sim T_{n-1}$ (c) Proportion: $Z = \frac{\hat{p}_1-\hat{p}_2-p_0}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$

(d) Variance: $F = \frac{s_1^2}{s_2^2}$. Multiple comparison: $\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$

Expected frequency: $E_{ij} = \frac{O_{i\cdot}\times O_{\cdot j}}{O_{\cdot\cdot}} = \frac{(row\ total)\times(column\ total)}{grand\ total}$

OTHER FORMULAS

Margin of error: $e = z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$

Pearson's coefficient of skewness: $(PS_k) = \frac{mean-mode}{standard\ deviation}$

Quartile coefficient of skewness: $(QS_k) = \frac{(Q_3-Q_2)-(Q_2-Q_1)}{Q_3-Q_1} = \frac{Q_3+Q_1-2Q_2}{Q_3-Q_1}$

Geometric Mean: $GM = \left(x_1^{f_1}.\ x_2^{f_2}.\ldots.x_n^{f_n}\right)^{1/N}$, *where* $N = \sum_{i=1}^{n} f_i$

Harmonic Mean: $HM = \frac{N}{\frac{1}{f_1}+\frac{1}{f_2}+\cdots+\frac{1}{f_n}}$

KAPASA MAKASA UNIVERSITY

SCHOOL OF APPLIED SCIENCE AND OPEN LEARNING

DEPARTMENT OF INFORMATION AND COMMUNICATION TECHNOLOGY

FINAL EXAMINATION

Course Title: **DISCRETE MATHEMATICS**

Course Code: **ITI 120**

Date: **October, 2025**

Duration: **3 HOURS**

Total Marks: **60**

GENERAL INSTRUCTIONS TO CANDIDATES:

1. There are **Six** (6) Questions in this Examination paper. Attempt any **Four** (4) Questions.
2. Show all essential working to earn full marks.
3. Each Question carries **15 Marks** and the total mark for the **Four** (4) attempted Questions is 60.
4. Start each answer on a new page.
5. Write clearly and readably.

Please Note: *Students that are caught with foreign material, copying or behave in any manner that is viewed as examination malpractice/misconduct will be subjected to a disciplinary hearing.*

QUESTION ONE

(a) Define the following terms

(i) Even and odd integers [2]

(ii) Prime and composite integers, [2]

(b) Given that $a = 2^2 \cdot 3 \cdot 5$ and $b = 2^4 \cdot 3^2$. Find

(i) $lcm\ (a,\ b)$, [2]

(ii) gcd $(a,\ b)$. [2]

(c) Assume that r and s are particular integers. Show that

(i) $4rs$ is an even integer, [2]

(ii) $6r + 4s^2 + 3$ is an odd integer, [2]

(iii) If r and s are both positive, then $r^2 + 2rs + s^2$ is composite. [3]

QUESTION TWO

(a) Evaluate the following

(i) $\left(\frac{20!}{18!}\right)^2$. [2]

(ii) 50! (use $n! = \sqrt{2\pi n} n^n e^{-n}$ (where $e \approx 2.718$) [2]

(iii) ${}_8C_3$ [1]

(b) Given: $A = \{1,2,3\}$ and $B = \{a, b\}$. Find

(i) $A \times B$ (ii) $B \times A$ (iii) $n(B \times B)$. [4]

(c) (i) A ball is drawn at random from a box containing 6 red balls, 4 white balls, and 5 blue balls. Determine the probability that it is red or white. [2]

(ii) Prove by mathematical induction that the sum of the first n positive integers is $\frac{1}{2}n(n+1)$, that is, $P(n) = 1 + 2 + 3 + \cdots + n = \frac{1}{2}n(n+1)$ [4]

QUESTION THREE

(a) Define the following terms

(i) a relation R, [2]

(ii) a function f. [2]

(b) Let $A = \{1,2,3,4,6\}$, and let R be the relation on A defined by " x divides y ", written $x \mid y$. (Note $x \mid y$ iff there exists an integer z such that $xz = y$.)

(i) Write R as a set of ordered pairs. [2]

(ii) Draw its directed graph. [3]

(iii) Find the inverse relation R^{-1} of R. [1]

(c) Solve the following equations

(i) $e^{2x} - 7e^{x} + 12 = 0$, [4]

(ii) $\log_5 x = -2$. [1]

QUESTION FOUR

(a) The density function of a random variable X is given by $f(x) = \frac{1}{2}x,\ 0 < x < 2$.

Evaluate $E(X)$ and $E(3X^2 - 2X)$. [5]

(b) Determine the domain and the range of the given function $f(x) = \sqrt{4 - x^2}$ [4]

(c) (i) Express the following in the form $\frac{a}{b}$ were a and b are integers and $b \neq 0$.

$0.\overline{56}$. [2]

(ii) Show that

$$\binom{17}{6} = \binom{16}{5} + \binom{16}{6}$$ [4]

QUESTION FIVE

(a) (i) Evaluate **43 *div* 9** and $\log_{16} 8$. [2]

(ii) Write the first two terms of the sequences defined by the formula

$a_n = \frac{n}{10 + n}$ for every integer $n > 3$, [1]

(iii) Prove by contradiction that $1 + 3\sqrt{2}$ is irrational. [3]

(b) Define a function $f: \mathbb{R} \times \mathbb{R} \to \mathbb{R} \times \mathbb{R}$ as follows: For all $(x, y) \in \mathbb{R} \times \mathbb{R}$,

$f(x, y) = (x + y, x - y)$. Show that f is one-to-one from $\mathbb{R} \times \mathbb{R}$ to itself. [3]

(c) Given $A = \{1,2,3,4\}$. Consider the following relation in A :

$$R = \{(1,1), (2,2), (2,3), (3,2), (4,2), (4,4)\}$$

(i) Draw its directed graph, [2]

(ii) find out whether or not R is reflexive, symmetric, and transitive. [4]

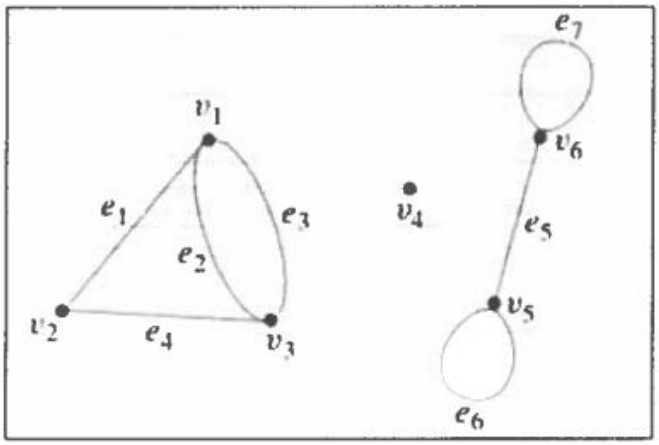
QUESTION SIX

(a) Prove that for any indexed class or any arbitrary collection of sets $\{A_i \mid i \in I\}$, we have

(i) $\left(\bigcup_{i\in I} A_i\right)^c = \bigcap_{i\in I} A_i^c$ [5]

(ii) $\left(\bigcap_{i\in I} A_i\right)^c = \bigcup_{i\in I} A_i^c$ [5]

(b) Consider the following graph below



(i) Write down the vertex set and the edge set, [2]

(ii) Give a table showing the edge-endpoint function. [2]

(iii) Find all parallel edges. [1]

END OF EXAMINATION

SCHOOL OF APPLIED SCIENCE AND OPEN LEARNING

DS 120: R-PROGRAMMING

2025 SESSIONAL EXAMINATION

DATE: 5th November 2025 **TIME: 3 hours** **TOTAL MARKS: 60 MARKS**

INSTRUCTIONS

1. Write your Student Identification Number on the Answer Booklet provided
2. **Answer all questions in Section A.**
3. **Answer any THREE (3) questions in Section B.**
4. Please write as clearly as possible, as illegible handwriting cannot be marked.
5. Number the answers to the questions clearly before answering

SECTION A: Each Question Carries 4 Marks

1. Which of the following is **not a data type in R**?
   a) Numeric
   b) Character
   c) Factor
   d) Boolean

2. What does the function `str()` in R do?
   a) Converts a variable to string
   b) Displays the structure of an object
   c) Splits a string into characters
   d) Concatenates strings

3. Which operator is used for **matrix multiplication** in R?
   a) *
   b) `%*%`
   c) `%%`
   d) :

4. Which function is used to **read CSV files** in R?
   a) `read.csv()`
   b) `read.table()`
   c) `readLines()`
   d) `readExcel()`

5. What is the output of the following R code?

```
x <- c(10, 20, 30, 40)
mean(x)
```

   a) 25
   b) 20
   c) 30
   d) 40

SECTION B: Each Question Carries 20 Marks

1. **Data Manipulation:**

I. Create a data frame called students with columns: Name, Age, Score.

II. Filter the students who scored more than 70.

III. Calculate the average score of all students. **[13.3 Marks]**

John Bend 18 20
Ruth Musa 20 25
Peter Mwanza 18 20
Mercy Hatyoka 18 60
Mulota Bwalya 20 70
Kelvin Chibwata 25 65

2. **Visualization:**

I. Using the mtcars dataset, create a **scatter plot** of mpg vs hp.

II. Color the points based on the number of cylinders (cyl). **[13.3 Marks]**

3. **Data Cleaning:**

I. Given a vector x <- c(10, NA, 20, NA, 30), replace all NA values with the **mean of non-NA values. [13.3 Marks]**

4. **Statistical Analysis:**

II. Using the iris dataset, calculate the **mean, median, and standard deviation** of Sepal.Length for each species. **[13.3 Marks]**

5. **Functions & Loops:**

I. Write a function in R that takes a numeric vector as input and returns a vector of **squared values.**

II. Use a for loop to print each element of the squared vector. **[13.3 Marks]**

Species	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
setosa	5.1	3.5	1.4	0.2
setosa	4.9	3	1.4	0.2
setosa	5	3.4	1.5	0.2
versicolor	6	2.9	4.5	1.5
versicolor	5.7	2.8	4.1	1.3
versicolor	6.3	3.3	4.7	1.6
virginica	6.5	3	5.2	2
virginica	7.1	3	5.9	2.1
virginica	6.9	3.1	5.4	2.1

Car_Model	mpg	wt hp	cyl
Mazda RX4	21	2.62	6
Datsun 710	22.8	2.32	4
Hornet 4 Drive	21.4	3.22	6
Hornet Sportabout	18.7	3.44	8
Valiant	18.1	3.46	6
Duster 360	14.3	3.57	8
Merc 240D	24.4	3.19	4
Merc 280	19.2	3.44	6
Toyota Corolla	33.9	1.84	4
Fiat 128	32.4	2.2	4

KAPASA MAKASA UNIVERSITY

SCHOOL OF APPLIED SCIENCE AND OPEN LEARNING

Department of Information and Communication Technology

FINAL EXAM

Course Title: Python Programing

Course Code: ITI 131

Date: 29TH JULY 2025

Duration: 3HRS

Total Marks: 100

Instructions to Candidates:

1. Answer questions 1 and 2 and any other 3 questions.
2. No unauthorized materials are allowed
3. No electronic devices allowed unless specified.
4. Start each answer on a new page.
5. Write clearly and neatly.

Please Note: *Students that are caught with foreign material, copying or behave in any manner that is viewed as examination malpractice/misconduct will be subjected to a disciplinary hearing.*

[QUESTION 1]

A. The code below shows a dictionary that indicates the foods that are available at your restaurant on a particular day. Come with a class of methods that have functions to display the menu, and other methods that allows the customer to pick what they want and calculate the total price of what they have picked. **[8 Marks]**

```
menue = {'nshima':'10','rice':'15', 'chicken': '22', 'soup':'7',
         'potatoes':'17','kapenta':'19','eggs':'15', 'vegiz':'11'}
```

B. Use list comprehension to write a single line of code that is going to store numbers from 1 to 50 with a step size of 5. Each of the numbers should be incremented by 4. **[2 Marks]**

C. The code below shows a calculator class with three methods to add, subtract and multiply. If tax is calculated at 37%, include two methods that would be able to help you calculate for both monthly tax and net pay in the class above so that the results would look as shown below. [7 Marks]

```
class Calculator:
    def __init__(self, num1,num2):
        self.num1 = num1
        self.num2 = num2
    def add(self):
        return self.num1+self.num2
    def sub(self):
        return self.num1-self.num2
    def mult(self):
        return self.num1*self.num2
```

```
tax=calct.get_tax(45000)
pay=calct.net_pay(45000)
print(f"tax is {tax}")
print(f"net pay is {pay}")

tax is 16875.0
net pay is 28125.0
```

D. If an employee has asked to contribute for pensions at a rate of 10% from his basic salary, write a method that will be part of the class above to calculate the deduction for pension. **[3 Marks]**

[QUESTION 2]

A. As an intern programmer, you have been asked to come up with python program that will be used for logins by students during registration. Use any methods learnt in class to create this login account that will ask a student for the user name and the password which are studentName and ****** respectively. If the student login is successful a message "Your login was successful, proceed with registration". If either the user name or the password is not correct a message "Your credentials are not correct, try again" should be displayed. If they decide to quit, they should end the program by submitting "quit" to the system.

[8 Marks]

B. The below snippet shows a list of students taking Python in 2022. Write the code using python that would add the name Ecilia to the list. **[2 Marks]**

```
python_students = ['Paul', 'Razeal','Chipo']
```

C. Write a function that builds a dictionary of three favorite courses of a student once called. The output after the function has been called could look like the one below. **[6 Marks]**

```
f_courses=courses('Python','Software Engineering','C++')

print(f_courses)

{'course1': 'Python', 'course2': 'Software Engineering', 'course3': 'C++'}
```

D. Explain what the code below will do when you run it. [4 Marks]

```
In [24]: courses = ['python','ICT Law', 'Databases','CS Foundations']

In [25]: for course in courses:
             if course == 'Databases':
                 print(course.title())
             else:
                 print(course.lower())
```

[QUESTION 3]

A. Describe what a Class is in Object oriented programming [3 Marks]
B. What is an instance of a Class in python programing? [2 Marks]
C. Come up with a program of your own depicting a Class, objects, methods and inheritance. [15 Marks]

[QUESTION 4]

A. A numpy array that was created with a python arange function has its output shown below. You are required to write a code that will generate another array **arr2** from arr1 that will be result in the second array as shown in the second output. [6 Marks]

```
print(arr1)

[0 3 6 9]

print(arr2)

[  0  27 216 729]
```

B. The numpy array shown below has two matrices created from a python arange function. Show the complete code and print the output so it matches what we have below. [4 Marks]

```
array([[[ 0,  1],
        [ 2,  3],
        [ 4,  5],
        [ 6,  7]],

       [[ 8,  9],
        [10, 11],
        [12, 13],
        [14, 15]]])
```

C. Use appropriate python code to create a numpy array of numbers ranging from 0 up to but not including 9 and reshape it accordingly. Use slicing to get the shown extract from the main array. Show code clearly with commenting. [6 Marks]

```
array([[3, 4],
       [6, 7]])
```

D. You have just started work at Fresher Construction Company after completing your first degree in Data Science and your Supervisor needs you to help them divide a 26 km stretch of the road into 7 equal points where they will put their Laybys. Use appropriate python code using numpy library for this task and display the specific points. [4 Marks] [2 Marks]

[QUESTION 5]

A. Create a numpy array of three random numbers and assign them to a variable x. After printing the array, it should appear like the first print. You are then required to convert the numpy array into a pandas Series and assign index to the random numbers as shown in the second display. The pandas Series is assigned to the variable names y. [6 Marks]

```
print(x)

[0.13724404 0.03521778 0.7135225 ]
```

```
print(y)

a    0.137244
b    0.035218
c    0.713522
dtype: float64
```

B. After the final Test, Data Science class obtained grades as follows: 12 of them got A+, 7 of them got A, 10 of them got B+, 3 of them got B and 9 of the got C. Create a dictionary that displays this information. Convert the dictionary into a pandas Series using appropriate code and display the result. From the displayed Series, slice the last three items and display them. Show all the code used clearly commented. [3 Marks]

C. Study a dataset provided below that was built with python. Use a python dictionary to come with the DataFrame that will display the dataset shown. [4 Marks]

frames

	Mark	Grace	Jackson	Jane
Pyhthon	20	17	92	54
Statistics	60	45	71	99
R_Programing	50	80	58	88
Language	30	62	48	81
Databases	90	52	65	70

[QUESTION 6]

A. Study the datasets DF1 and DF2 shown below. Build the DataFrames from the Dictionary then use appropriate python code merge the two DataFrames to an equivalent right join on the 'id' column. [4 Marks]

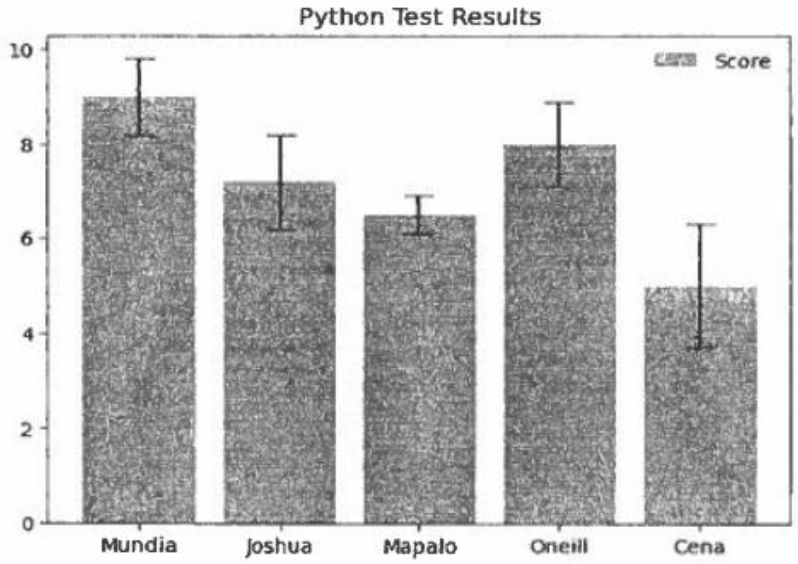
DF1

	id	price	location
0	ball	12.33	tsk
1	pencil	11.44	ktw
2	pen	33.21	lgst
3	mug	13.21	kbw
4	rubber	33.62	mnsa

DF2

	id	color	stknum
0	penci	white	2
1	pencil	red	4
2	ball	red	8
3	pen	black	6

B. The bar chart below shows the performance of 5 students in a python programming Test two. Use Matplotlib to produce the bar chart with capsize and alpha values of 7 each as shown below. [4 Marks]



C. Use the data above to produce a pie chart to represent the same information.
[2 Marks]

The End. The future is in your Hands