- Attempt ALL questions. Each question carries **6 points. Each part (a, b, and c) of the question is of 2 points.** There are **05 questions** written on 2 pages.
- At the end of the exam, please return the question paper along with the answer script.
- You may make documented assumptions, without contradicting the crux of the question.
- Please answer all the questions according to the sequence given in the question paper.

Time Allowed: 120 minutes (i.e. 2 hours)   **Total Points: 30**

Question # 1

a) What is Data science? Why statistics is essential for data science? Differentiate between Descriptive and Inferential Statistics.

b) Define the following terms in reference to Statistics: Population, Sample, Variable, Event and Sample Space.

c) Briefly explain the terms: Box plot and Inter quartile range (IQR) .

Question # 2

a) The probability that a student will pass English is $\frac{3}{8}$ and the probability that he will pass Mathematics is $\frac{3}{4}$. If the probability that he will pass either one subject is $\frac{7}{8}$, what is the probability that he:

   i.   passes both subjects   ii. fails English

   iii.   passes English but fails Mathematics   iv. fails both subjects

b) Show that the function $g(x) = \begin{cases} \frac{x}{6} & if\ x = 1, 2, 3 \\ 0 & otherwise \end{cases}$ is a probability mass function of a random variable x? Show the probability distribution for x and depict it by a graph.

c) For the function given in question "2(b)", solve the following:

   i. $P(x = 2)$ ii. $P(x \leq 2)$ iii. $P(x < 2)$ iv. $P\ (x \leq 3)$

Question # 3

In a survey, conducted by National Center for Health Statistics (NCHS), the sample mean height of women in the United State (ages 20-29) was 64.30 inches, with a standard deviation of 2.62 inches. Estimate the percent of women whose heights are:

a) Between 59.06 inches and 64.30 inches
b) Between 61.68 inches and 69.54 inches
c) Show the results of part a) and b) with shades in Bell-Shaped Distribution diagram.

Question # 4

A study investigated causes of sudden deaths in western region of Paris and France. A sample of 523 such deaths revealed the following:

	Cardiovascular	Cerebral	Respiratory	Others	Total
Males	264	38	36	21	359
Females	89	27	29	19	164
Total	353	65	65	40	523

Suppose one of these cases is randomly selected.

a) What is the probability the person was female? Given the cause was cardiovascular in nature, what is the probability the person was female?
b) Given the person was female, what is the probability the cause was cardiovascular in nature?
c) Given the cause was cerebral or respiratory in nature, what is the probability the person was male?

Question # 5

There is a screening test for prostate cancer that looks at the level of PSA in the blood. There are a number of reasons besides cancer that a man can have elevated PSA levels. In addition, many types of cancer develop so slowly that that they are never a problem. Unfortunately, there is currently no test to distinguish the different types and using the test is controversial because it is hard to quantify the accuracy rates and the harm done by false positives.

For this problem we'll call a positive test a true positive if it catches a dangerous type of cancer. We'll assume the following numbers:

Rate of cancer among men over 50 = 0.0005
True positive rate for the test = 0.9
False positive rate for the test = 0.01

Let T be the event a man has a positive test and let D be the event a man has a dangerous type of the disease. Find the probability that a man does have dangerous type of disease given that the tests are positive and negative.