SECTION A

1. Consider the following data which represent the lives of 20 similar car batteries recorded to the nearest tenth of a year. The batteries are guaranteed to last 3 years by its manufacturer.

   4.1, 2.6, 1.9, 2.2, 2.9, 3.4, 4.7, 1.6, 3.8, 3.5, 3.2, 3.4, 3.4, 3.5, 3.1, 4.3, 3.3, 3.1, 2.5, 2.6.

   (a) Display the above data as a stem-and-leaf plot. Describe the shape of the distribution. [5 marks]

   (b) Calculate the median and the interquartile range of the above data. [3 marks]

2. (a) Keep on tossing a biased coin until you get a head. If $P(H) = 0.4$ for this coin, what is the probability that you will stop at the 4-th toss? [3 marks]

   (b) (i) Let A and B be two mutually exclusive events with $P(A) > 0$ and $P(B) > 0$. Are they independent? Give reason. [1 mark]

   (ii) Let A and B be two mutually exclusive events with $P(A) = 0$ and $P(B) > 0$. Are they independent? Give reason. [1 mark]

   (iii) Let A and B be two events such that $A \subset B$, $P(A) > 0$ and $P(B) < 1$. Are A and B mutually exclusive? Are they independent? Give reason. [3 marks]

3. Polygraph tests (lie-detectors) are routinely administered to employees in sensitive positions. Let $(+)$ denote the event that the polygraph reading is positive and suggests that the subject is lying, $(-)$ denote the event that the polygraph reading is negative and suggests that the subject is not lying, L denote the event that the subject is lying and T denote the event that the subject is telling the truth. According to studies of polygraph reliability,

$$P(+|L) = 0.88 \text{ and } P(-|T) = 0.86.$$

   Suppose that for a particular question, the vast majority of subjects have no reason to lie so that $P(L) = 0.01$ and $P(T) = 0.99$.

   (a) What is the probability of a positive response on the polygraph? [3 marks]

   (b) Given that a person produces a positive response on the polygraph, what is the probability that she is in fact telling the truth? [5 marks]

4. Let X be a discrete random variable taking values $0, 1, \ldots, 4$ with probability mass function $f(x) = P(X = x)$ given by

$$f(x) = c(x^2 + 4), \quad \text{for } x = 0,\ 1,\ 2,\ 3,\ 4,$$

for some constant c.

(i) Find the value of c. [1 mark]

(ii) Calculate $P(X > 0)$. [1 mark]

(iii) Calculate $P(1 < X < 4)$. [1 mark]

(iv) Calculate the conditional probability $P(X > 1 | X > 0)$. [3 marks]

(v) Calculate $E[X]$. [2 marks]

5. Each of the following statements from 5(a) to 5(h) (altogether eight) is either true or false. Answer them by writing either 'True' or 'False' against the statement number in your answer script. For each statement, the correct answer is awarded 1 mark.

(a) The probability mass function (pmf) of a discrete random variable must always be less than or equal to 1.

(b) The probability density function (pdf) of a continuous random variable must always be less than or equal to 1.

(c) The CLT (central limit theorem) states that when the sample size n is large, the population distribution is approximately normal.

(d) If we choose a large sample of size $n = 10000$ from a Poisson distribution with mean 1 and variance 1 and then draw a histogram, the histogram should look like a normal density with mean 1 and variance 1.

(e) If we choose a large sample of size $n = 10000$ from a normal population with mean 0 and variance 1 and then draw a histogram, the histogram should look like a normal density with mean 0 and variance $1/n = 0.0001$.

(f) A confidence interval for the sample mean does not make sense.

(g) Based on the same sample, a 99% confidence interval for the mean will be always wider than a 95% confidence interval.

(h) Suppose that a 95% confidence interval for the population mean μ turns out to be $(0.50, 0.59)$. A test for the null hypothesis $\mu = 0.49$ based on the same data will then be rejected at 5% level.

SECTION B

6. Let X be a random variable having the probability density function (pdf)

$$f(x) = \begin{cases} c\,(1-x) & \text{if } 0 \leq x \leq 1, \\ 0 & \text{otherwise,} \end{cases}$$

where $c > 0$ is a constant.

(a) Show that $c = 2$. [3 marks]

(b) Compute the mean and the variance of X. [8 marks]

(c) Compute $P[X < 0.5]$. [3 marks]

(d) A dealer's profit, in units of 1000 pounds, on selling a new automobile is given by X^2, where X is as above. Find the probability that the profit will be less than 250 pounds on the next automobile sold by the dealership. [6 marks]

7. Exam Scores in modules with large number of students often have normal distribution. Suppose that in one such module, the average score was 74 and the standard deviation was 7.

(a) If the pass mark was set at 60, what proportion of students passed the examination? [3 marks]

(b) If the top 12% of the students were given A's, what is the lowest possible score of a student who received an A? [8 marks]

(c) Find the conditional probability that a randomly chosen student has got 'A' given that he has passed the examination. [9 marks]

8. The Poisson distribution is useful for describing rare random events such as severe storms. In the 98-year period from 1900-1997, there were 159 U.S. landfallng hurricanes and the distribution of the number of landfalling hurricanes/year is given in the table below.

Number of hurricanes/year	0	1	2	3	4	5	6
Frequency	18	34	24	16	3	1	2

(a) Use the data to estimate the mean number of landfalling hurricanes/year. [3 marks]

(b) Hence compute the expected frequencies for the different number of landfalling hurricanes/year. [8 marks]

(c) Use a goodness of fit test at 5% level to check whether the number of landfalling hurricanes/year follows a Poisson distribution. [9 marks]

9. Suppose that the diastolic blood pressure X in a large population of hypertensive women has a normal distribution with mean 100 mmHg and standard deviation 16 mmHg.

(a) Find the probability that a randomly chosen woman from this population will have blood pressure more than 132 mmHg. [3 marks]

Give numerical answers for each of the following probabilities in (b), (c) and (e) using the most suitable approximation.

(b) Probability that there are exactly 2 women with blood pressure more than 132 mmHg out of a random sample of 100 hypertensive women. [5 marks]

(c) Probability that the average blood pressure of those 100 hypertensive women selected will be more than 104 mmHg. [5 marks]

(d) Suppose now 300 independent samples, each of size 100 were taken from this population with replacement (that is, each sample was returned to the population before taking a new sample). Let $\{\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{300}\}$ denote 300 sample means. If we draw a histogram of these 300 means, describe in words what distribution/density it would look like? You don't need to draw any picture. [2 marks]

(e) Probability that there are at most 2 means out of these 300 sample means $\{\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{300}\}$ for which the average blood pressure will be more than 104 mmHg. [5 marks]

10. (a) Suppose that in a city the number of suicides can be approximated by a Poisson process with rate $\lambda = 0.33$ per month. What is the probability of two suicides in one week ? (Assume that one month consists of 4 weeks). [6 marks]

(b) It is known that under certain growing conditions, the heights of a certain type of plant are normally distributed with variance 1.38 cm^2. A sample of 250 plants give mean height of 23.6 cm.

i. Calculate a 95% confidence interval for the mean height of such plants. [5 marks]

ii. Explain the meanning of the confidence interval. [2 marks]

iii. Test at 1% level whether the mean height of plants is different from 23.5 cm and state your conclusion. [6 marks]

iv. Compare your result with the confidence interval in (i) and make some comments. [1 mark]