This examination contains two sections: Section A and Section B. Questions in Section A are worth 5 marks each and questions in Section B are worth 20 marks each. Candidates should attempt all of the questions in Section A and two questions from the three in Section B. [New] Cambridge Statistical Tables are to be provided.
Section A

1. Suppose $A$, $B$, and $C$, are events such that

\[ P(A \cap B) = \frac{3}{10}, \quad P(B \cap C) = \frac{2}{10}, \quad P(A \cap C) = \frac{1}{2} \]

\[ P(A \cap B \cap C) = \frac{1}{10}, \quad P(C) = \frac{6}{10}. \]

Further suppose that $A$ is independent of $B$, and $B$ is independent of $C$.

(a) Show that $P(B) = \frac{1}{3}$. \[1\]
(b) Find the value of $P(A)$. \[1\]
(c) With justification, determine whether the inequality

\[ P(A|C) < P(B|C) \]

is true or false. \[3\]

2. It is believed that there is a 30% chance that an adult in a certain district will contract the illness known as influenza during the winter, independently of everyone else. Let the r.v. $Y$ represent the number of people who contract the illness in a random sample of 10 adults.

(a) State, as precisely as possible, the distribution of $Y$. \[2\]
(b) Using tables, or otherwise, determine the exact probability that more than 2 people in a random sample of 10 adults contract the illness. \[3\]
3. Suppose that the number of cars that pass through a certain roundabout between 1 a.m. and 2 a.m. can be represented by a r.v. $X$ that has a Poisson distribution with mean value 15, i.e.

$$p_X(n) = e^{-15} \frac{(15)^n}{n!} \quad n = 0, 1, 2, \ldots$$

(a) What is the probability that no more than 10 cars pass through the roundabout in that time period? [2]

(b) Let $m$ be a non-negative integer. Find the smallest value of $m$ such that

$$P(X \geq m) < 0.05$$

[You may use tables to facilitate your calculation].

4. Suppose that the r.v. $X$ has moment generating function (m.g.f.)

$$M_X(t) = \frac{e^{\alpha t}}{1 - \beta^2 t^2}$$

for suitably defined $t$.

(a) Show that the expected value of $X$, i.e. $E[X]$, is equal to $\alpha$. [2]

(b) Further suppose that $Y = 3X + 4$ and $\alpha = \beta = 1/3$. Deduce an explicit expression for the m.g.f. of $Y$ as a function of $t$ only. [3]

5. Suppose that the discrete r.v. $Y$ has a probability generating function given by

$$G_Y(z) = \left[ \frac{\theta}{1 - (1 - \theta)z} \right]^\alpha$$

where $\alpha > 0$ and $0 < \theta < 1$, and for suitable $z$.

(a) Show that

$$E[Y] = \frac{\alpha(1 - \theta)}{\theta} \quad \text{and} \quad E[Y^2] = \frac{\alpha^2(1 - \theta)^2}{\theta^2} + \frac{\alpha(1 - \theta)}{\theta^2}.$$  [3]

(b) Find the value of the variance of $Y$, i.e. var($Y$), when $\alpha = 5$ and $\theta = 1/4$. [2]
6. Suppose that the random variables $X$ and $Y$ have a joint probability density function of the form

$$f_{(X,Y)}(x, y) = \begin{cases} kx^4 & \text{for } x > 0, y > 0, x + y < 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) Sketch the joint range space of $X$ and $Y$, i.e. $R_{(X,Y)}$, [1]
(b) Deduce the value of $k$. [2]
(c) What is the expected value of $X$, i.e. $E[X]$? [2]

7. A coffee machine is calibrated so that the amount of liquid dispensed in each cup behaves like a random variable with mean 150 millilitres (ml) with a standard deviation of 12 ml.

Using the central limit theorem, determine, as accurately and precisely as possible, the approximate probability that the average (i.e. sample mean) amount dispensed in a random sample of 64 cups is greater than or equal to 148 ml but less than or equal to 153 ml. [5]

8. A random sample of 423 residents in a particular borough were asked whether they preferred the local play area to be converted into a cricket pitch, or a basketball facility.

Data on the responses are summarized below and are classified according to age range and preference of the respondents.

<table>
<thead>
<tr>
<th>Age (in years)</th>
<th>Preference</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;20</td>
<td>Cricket</td>
</tr>
<tr>
<td>20-35</td>
<td>33</td>
</tr>
<tr>
<td>&gt;35</td>
<td>41</td>
</tr>
<tr>
<td>33</td>
<td>60</td>
</tr>
<tr>
<td>55</td>
<td>30</td>
</tr>
<tr>
<td>204</td>
<td></td>
</tr>
</tbody>
</table>

Using an appropriate test, determine whether there is any evidence to reject the hypothesis of independence between preferences of the respondents and age group at the 5% level of significance. [5]
9. A manufacturer of rechargeable electric toothbrushes is trying to decide whether to install a different brand of battery in its appliances in order to extend the time until recharging is needed.

Two independent random samples, which are assumed to be Normally distributed, were drawn from the brand of battery currently used by the manufacturer, and the proposed new brand. These batteries were fully 'charged-up' and left to run under appropriate experimental conditions. The lifetimes (in hours) of the batteries are presented below:

A: Current Brand

| 7.51 | 7.33 | 7.92 | 7.63 | 7.68 | 7.65 | 7.28 | 7.52 |

B: Proposed New Brand

| 7.92 | 7.69 | 7.65 | 7.61 | 7.84 | 7.84 |

(a) It is believed that the variances of the lifetimes of the two brands of battery are the same. Determine whether there is any evidence against this hypothesis at the 5% level of significance. [10]

(b) Assuming that the variances across the two brands of battery are the same, test whether the proposed new brand has a longer lifetime than the current brand at both the 5% and 1% levels of significance. [10]

[You are cautioned to carry as many significant figures in your calculations as possible and to only round up the final answers that will used to address each part of the question].

Please turn over
10. A driving school is in the process of drawing up a new performance related pay scale for its driving instructors. To ensure that this is fair, the school would like to assess the variability of the number of hours of tuition required by pupils prior to taking the practical car test.

A random sample of 11 students who recently completed their practical driving tests were drawn from its database and the number of hours of tuition for each student was calculated. The data are presented below.

28.89 30.97 29.47 28.84 26.89
33.71 32.36 31.60 32.91 25.68 29.66

It is assumed that the data are Normally distributed, and that the population from which the data were drawn has an unknown mean \( \mu \) and unknown variance \( \sigma^2 \).

(a) With justification, suggest an appropriate point estimator for \( \sigma^2 \), and calculate its observed value for the data above. [4]

(b) Suppose the sample size is \( n \), and that the sample variance is given by \( s^2 \). What is the distribution of \( (n - 1)s^2/\sigma^2 \)? [2]

(c) Calculate a 90% confidence interval for \( \sigma^2 \). [4]

(d) Show that the expectation of the length of a 100(1 - \( \alpha \))% confidence interval for \( \sigma^2 \), based on a sample of size \( n \), is given by the formula

\[
(n - 1)\sigma^2 \left[ \frac{1}{\chi^2_{1 - \alpha/2, n-1}} - \frac{1}{\chi^2_{\alpha/2, n-1}} \right].
\]

[e] Determine the smallest sample size \( n \) in the range 13 to 16 inclusive, i.e. for \( n \in \{13, 14, 15, 16\} \), that would ensure that the expected length of a 95% confidence interval for \( \sigma^2 \) is less than or equal to 2\( \sigma^2 \). [5]
11. A member of parliament (MP) wants to introduce some legislation under a *private members’ bills procedure* in an attempt to get public service television broadcasters to only show ‘educational’ programmes during the late afternoon and evening. In order to make this case, the MP conducts a survey on a random sample of 12 secondary school children in her constituency. Each child was asked to keep a diary of the times when he/she was watching television during a particular week. These data were used to ascertain the number of hours of television each child watched during that week, which are given below.

\[
12.68, 11.70, 11.66, 10.48, 9.61, 11.26, \\
10.49, 10.79, 10.39, 9.39, 10.22, 9.92
\]

It is assumed that these data are Normally distributed, with unknown parameters \( \mu \) and \( \sigma^2 \), which represent the mean and variance, respectively.

(a) It is widely believed that children who attend secondary school watch about 10 hours of television per week. By constructing an appropriate confidence interval, and not otherwise, test this claim at the 5% level of significance. \([10]\]

(b) In order to strengthen her arguments, the MP would like to claim that children watch more than ten and a quarter (10.25) hours of television per week. By constructing and carrying out appropriate tests at both the 5% and 1% levels of significance, comment on whether she can make this claim or not. \([10]\)