

Section B: Probability and Statistics [60 marks]

- 5** For events A and B , it is given that $P(A \cup B) = 0.8$ and $P(A|B) = 0.5$. It is also given that events A and B are independent.

(i) Find $P(B)$. [2]

A third event C is such that events A and C are independent and events B and C are independent.

(ii) Given also that $P(C) = 0.5$, find exactly the maximum and minimum possible values of $P(A \cap B \cap C)$. [3]

- 6** Drew has a sock drawer which contains 3 pairs of red socks, 2 pairs of white socks and 1 pair of black socks. In the mornings, Drew randomly pulls out socks from the drawer, one sock at a time, until he has a pair of socks which matched in colour. The random variable S denote the total number of socks Drew pulls out in a morning.

(i) Determine the probability distribution of S . [3]

(ii) Find $E(S)$ and $\text{Var}(S)$. [2]

(iii) Find the probability that Drew took more than 2 pulls to obtain a pair of matching socks for 7 consecutive mornings. [1]

- 7** Peter and Calvin are brothers. They went to a birthday party with 6 other friends.

(i) The 8 friends started off the party with dinner, sitting down on 8 chairs around a circular dining table. Given that there was at least one other person seated between the two brothers, find the number of different arrangements of the 8 friends at the table. [2]

(ii) The 8 friends then decided to play a game of musical chairs. 5 chairs were placed in one row. When the music stopped, Peter could not get a seat but Calvin managed to sit on the chair in the middle of the row. Find the number of ways that the 5 chairs could be occupied. [1]

(iii) The 8 friends decided to watch a movie to end off their night. There are 4 movies for selection. Each person will vote for a movie to watch, and the movie with the highest number of votes will be played. Find the total number of ways the 8 friends could have voted such that there were movies tied for the highest number of votes. [4]

- 8 Wally is looking for an investment plan that generates a monthly income of more than \$200. He shortlisted an investment plan whose prospectus promises investors a variable monthly income with a mean of \$200 per month. Wally decides to research further by collecting a random sample of 36 of the plan's past monthly income returns. The monthly income returns, \$ x , are summarised as follows.

$$\sum (x - 200) = 662.4 \quad \sum (x - 200)^2 = 307141.56$$

- (i) Test, at the 10% level of significance, whether Wally should invest in this plan. You should state your hypotheses and define any symbols you use. [6]
- (ii) Explain what is meant by the phrase '10% significance level' in the context of this question. [1]
- 9 The Bacterial Filtration Efficiency (BFE), in percentages, of Brand *A* surgical masks against the *E. Coli* bacteria is assumed to have the distribution $N(\mu, \sigma^2)$. It is given that the proportion of surgical masks having a BFE less than 95.70% is the same as the proportion of the surgical masks having a BFE more than 95.78%.

- (i) The probability that the BFE of a Brand *A* surgical mask is more than 95.78% is 0.0912. Find the values of μ and σ . Leave your answers to 2 decimal places. [3]

The BFE of Brand *A* surgical masks against the *S. Aureus* bacteria have the distribution $N(91.09, 0.08^2)$ while the BFE of Brand *B* surgical masks against the *S. Aureus* bacteria have the distribution $N(92.19, 0.03^2)$.

- (ii) Given that the probability that the BFE of a randomly chosen Brand *B* surgical mask differs from the sample mean BFE of n randomly chosen Brand *A* surgical masks by at most 1.15% is at least 0.9405, find the least value of n . [4]

The masses in grams of a box of Brand *A* surgical masks have the distribution $N(203, \sigma_1^2)$ while the masses in grams of a box of Brand *B* surgical masks have the distribution $N(203, \sigma_2^2)$.

- (iii) Find the probability that three times the mass of a randomly chosen box of Brand *A* surgical masks exceeds the total mass of 3 randomly chosen boxes of Brand *B* surgical masks. You should state the parameters of any distributions that you use. [2]

- 10** A popular investment manager claims that the mean dividend yield of his clients is 12%. A research firm investigates further by collecting a random sample of 40 of the manager's clients. The sample was found to have a mean dividend yield of 10.9% and a standard deviation of 2.5%.

(i) State, giving a reason, whether there is a need to make any assumption about the distribution of the dividend yield earned by the manager's clients in order for a hypothesis test to be valid. [1]

(ii) Test, at the 5% level of significance, whether the manager overstated his claim. [5]

It was later known that the standard deviation of the dividend yield of the popular investment manager's clients is 2.3%. A potential investor collected the dividend yield data of a small sample size, n , of the manager's clients. The sample was found to have a mean dividend yield of 10.2%.

(iii) State, in context, two assumptions that are needed for you to carry out a test to examine the manager's claim that the mean dividend yield of his clients is 12%. [2]

(iv) Hence find the least value of n such that the manager's claim, at the 5% level of significance, will be rejected. You should state your hypotheses clearly. [4]

- 11** In an epidemic outbreak, an infectious disease spread through a population. It is estimated that 5% of the population are infected. A quick test for the presence of the disease was carried out on the population. An infected person tested positive 90% of the time, while a non-infected person tested positive 6% of the time.

(i) Draw a probability tree diagram to represent the above scenario and find the probability that the quick test is accurate. [2]

(ii) Find the probability that a person is infected given that he tested positive. [2]

(iii) To increase the accuracy of identifying infected or non-infected people, a more comprehensive but more costly laboratory confirmation test can be carried out.

Give **one** reason why you would recommend that people who tested positive on the quick test be sent for the laboratory confirmation test. Justify your answer. [1]

State also, **one** drawback of selecting only the people who tested positive on the quick test to be sent for the laboratory confirmation test. [1]

In another population, 5% of the population were confirmed to be infected. A sample of 20 people were randomly chosen from this population.

(iv) State one assumption needed for the number of infected people out of these 20 people to be well-modelled by a binomial distribution, and hence find the probability that there were fewer than 2 infected people out of these 20 people. [2]

(v) Find the probability that, for the 20 people chosen, there were more than 2 infected people and the first infected person was the tenth person chosen for the sample. [3]

- (vi) A total of n random samples of 20 people were chosen from this population. Given that the probability of at most 5 of the samples having fewer than 2 infected people is greater than 0.1, find the largest possible value of n . [3]