

**Paper Reference 9ST0/01**  
**Pearson Edexcel Level 3 GCE**

**Statistics**  
**Advanced**  
**PAPER 1: Data and Probability**

**Time: 2 hours**

**YOU MUST HAVE**

**Statistical formulae and tables booklet**  
**Calculator**

**YOU WILL BE GIVEN**

**Answer Booklet**  
**Data Booklet**

**Y69600RA**

**Candidates may use any calculator allowed by Pearson regulations. Calculators must not have retrievable mathematical formulae stored in them.**

## **INSTRUCTIONS**

**In the boxes on the Answer Booklet and on the Data Booklet, write your name, centre number and candidate number.**

**Answer ALL questions and ensure that your answers to parts of questions are clearly labelled.**

**Answer the questions in the Answer Booklet – there may be more space than you need.**

**You should show sufficient working to make your methods clear. Answers without working may not gain full credit.**

**Unless otherwise stated, inexact answers should be given to three significant figures.**

**Unless otherwise stated, statistical tests should be carried out at the 5% significance level.**

**Turn over**

## **INFORMATION**

**A booklet ‘Statistical formulae and tables’ is provided.**

**There are 7 questions in this Question Paper. The total mark for this paper is 80**

**The marks for EACH question are shown in brackets – use this as a guide as to how much time to spend on each question.**

**Turn over**

## **ADVICE**

**Read each question carefully before you start to answer it.**

**Try to answer every question.**

**Check your answers if you have time at the end.**

**If you change your mind about an answer, cross it out and put your new answer and any working underneath.**

---

**Answer ALL questions.  
Write your answers in the  
Answer Booklet.**

- 1. Refer to Diagram 1 and Diagram 2 for Question 1 in the Data Booklet.**

**The journey times of commuters for a selection of those travelling into two London train stations are given in Diagram 1 and Diagram 2 in the Data Booklet.**

**(continued on the next page)**

**1. continued.**

**These times are the time taken from  
LEAVING THEIR HOUSE to arriving  
at their destination station.**

**(a) Compare the average and spread  
of journey times for commuters  
travelling to these two stations.  
(6 marks)**

**(continued on the next page)**

**Turn over**



**1. continued.**

**These journeys are made up of two parts.**

**The first part of the total journey is the journey travelling from the commuter's home to their local station, where they take the train to London King's Cross or London Waterloo.**

**This part of the journey may include waiting for this train to arrive.**

**(continued on the next page)**

**1. continued.**

**The second part of the total journey is the journey by train to London King's Cross or London Waterloo.**

**All passengers in Diagram 1 in the Data Booklet took the 7:15 train from Stevenage to London King's Cross.**

**All passengers in Diagram 2 in the Data Booklet took the 7:45 train from Wimbledon to London Waterloo.**

**(continued on the next page)**

**Turn over**

**1. continued.**

**All the data is from the same day.**

**Lyra looks at the information in Diagram 1 and Diagram 2 in the Data Booklet and concludes that the train journey time from Wimbledon to London Waterloo is shorter than that from Stevenage to London King's Cross.**

**(b) State an assumption that Lyra has made in order for this conclusion to be valid.**

**(1 mark)**

**(continued on the next page)**

**Turn over**

**1. continued.**

- (c) Identify TWO sources of variability for these commuters from Wimbledon to London Waterloo?  
(2 marks)**

**(Total for Question 1 is 9 marks)**

---

- 2. Refer to the table and the diagram for Question 2 in the Data Booklet.**

**Alana is researching show dogs.**

**Show dogs have a desired range of heights.**

**The maximum and minimum desired male and female heights, in inches, for a random sample of breeds of show dog are shown in the table in the Data Booklet.**

**(continued on the next page)**

**Turn over**

**2. continued.**

**Alana plots the minimum desired male heights and minimum desired female heights in the graph in the Data Booklet.**

**(a) Explain why Alana might have chosen to plot her data before calculating a correlation coefficient.**

**(2 marks)**

**(continued on the next page)**

**2. continued.**

**After examining the graph in the Data Booklet Alana realises that one of the dog breeds has had the male and female columns recorded the wrong way round.**

**(b) State, with a reason, which breed you believe has been recorded incorrectly.**

**(2 marks)**

**(continued on the next page)**

**Turn over**

**2. continued.**

**The PMCC between the minimum desired height for males and females was calculated as  $r = 0.988$**

**Alana says that the PMCC will be higher if the heights were measured in centimetres.**

**(c) Explain why Alana is incorrect.  
(1 mark)**

**(continued on the next page)**



**2. continued.**

**Alana edits the data, swapping the values she realised had been recorded incorrectly.**

- (d) How would you expect the PMCC of her corrected data to compare to the PMCC of the original data,  $r = 0.988$ ?  
(1 mark)**

**(continued on the next page)**

**2. continued.**

**Alana states that it would be impossible for there to be a negative correlation between the minimum desired female height and maximum desired female height for all breeds of show dog.**

**(e) Discuss Alana's statement.**

**You should include reasoned comments on whether or not you believe she is correct.**

**(4 marks)**

**(Total for Question 2 is 10 marks)**

---

**Turn over**

3. In 2011 Boston College psychologists, Angelina Hawley–Dolan and Ellen Winner, investigated whether paintings of professional abstract artists could be differentiated from those of a five–year–old child.

Leo, a psychology student, decides to do his own version of this experiment.

(continued on the next page)

**3. continued.**

**Suggest how Leo might set up his experiment.**

**You should include comments about**

- **how he will source the data**
- **the data he should collect**
- **how he could minimise bias.**

**(Total for Question 3 is 7 marks)**

---

**4. Shai has created a video game.**

**She makes her game free to play, with players able to buy gems to spend in the game for bonuses.**

**She creates a model to predict the income that she can expect from her game.**

**Shai's data shows that the majority of players do not spend any money on gems.**

**(continued on the next page)**

**4. continued.**

**A small number of players, known as ‘whales’, make up the majority of the revenue for games by buying the gems.**

- (a) Give a reason why the normal distribution is unlikely to be a suitable model for the amount a randomly selected player spends in a given month.**
- (1 mark)**

**(continued on the next page)**

**Turn over**

**4. continued.**

**Shai instead chooses to model the spend, in pounds, of a randomly selected whale in a given month.**

**She uses the distribution**

$$\mathbf{W \sim N(50, 100)}$$

**She assumes that the amount a whale spends each month is independent from their spend in previous months.**

**(b) Find  $P(W < 40)$**

**(1 mark)**

**(continued on the next page)**

**Turn over**

**4. continued.**

**(c) Using Shai's model, calculate the probability that a randomly selected whale has a mean spend of less than £40 a month over a year.**

**(2 marks)**

**(d) Explain why the probability calculated in part (c) MUST be smaller than the probability calculated in part (b)**

**(1 mark)**

**(continued on the next page)**

**Turn over**



4. continued.

(e) Explain why  $P(W \leq 40)$  and  $P(W < 40)$  are equal.

(1 mark)

(f) Calculate the probability that a randomly selected whale spends **MORE** than **£40** in each of four consecutive months.

(2 marks)

(continued on the next page)

**4. continued.**

**(g) Calculate the probability that a randomly selected whale spends more than £40 in at least 8 of the months in a one-year period.**

**(3 marks)**

**(h) Calculate the probability that a randomly selected whale spends less than £40 in a month that they spent more than £30**

**(3 marks)**

**(continued on the next page)**

**Turn over**

**4. continued.**

**In Shai's game, a bundle of gems costs £5**

**The game also offers unlimited gems for any player spending £60 a month.**

- (i) Discuss the appropriateness or otherwise of the normal distribution to model the spend of a whale in light of this information.**
- (3 marks)**

**(Total for Question 4 is 17 marks)**

---

**Turn over**

- 5. Tom owns a busy cinema with two screens.**

**Screen 1 has 120 seats, and  
Screen 2 has 300 seats.**

**He shows a different film on each screen.**

**Children and pensioners are eligible for a concession ticket that is cheaper than a standard ticket.**

**From past experience Tom estimates that the proportion of concession tickets sold for the most popular 7pm showing for the film on Screen 1 is 35% and the proportion for the film on Screen 2 is 15%**

**(continued on the next page)**

**Turn over**

**5. continued.**

**Both screens show films at 7pm  
one night.**

**The random variable  $X$  denotes  
the number of concession tickets  
purchased for the 7pm showing of  
the film on SCREEN 1**

**The random variable  $Y$  denotes  
the number of concession tickets  
purchased for the 7pm showing of  
the film on SCREEN 2**

**(continued on the next page)**

**Turn over**

**5. continued.**

**The random variable  $T$  denotes the TOTAL number of concessions tickets purchased for BOTH showings of films at 7pm that night.**

**Tom assumes that all the tickets will be purchased, and that the probability of a concession ticket being purchased is independent from one customer to the next.**

**(continued on the next page)**

**5. continued.**

**(a) By modelling each of  $X$  and  $Y$  as an independent binomial distribution,**

**(i) show that the distribution of  $T$  can be approximated as  $N(87, 65 \cdot 55)$**

**(4 marks)**

**(ii) estimate the probability that at least **100** concessions tickets are purchased in total.**

**(2 marks)**

**(continued on the next page)**

**Turn over**

**5. continued.**

**(b) Assuming that it is valid to model  $X$  and  $Y$  as binomial distributions, give TWO reasons to justify the use of the APPROXIMATION in part (a) (2 marks)**

**(c) Give TWO reasons in context why the assumptions that Tom made may NOT be justified. (2 marks)**

**(continued on the next page)**

**Turn over**



**5. continued.**

**Tom charges £4 for a concession ticket, and £6 for a regular ticket.**

**It costs him £200, in total, to run the cinema for the 7pm showing.**

**(d) Assuming that all tickets are sold for both screens,**

- (i) write an expression for the number of regular tickets sold in terms of  $T$**   
**(1 mark)**

**(continued on the next page)**

**Turn over**

**5. (d) continued.**

**(ii) find the probability Tom  
earns at least £2 125 of  
profit.**

**Note that profit is total  
income from sales minus  
costs.**

**(5 marks)**

**(Total for Question 5 is 16 marks)**

---

**Turn over**

**6. Refer to the information for Question 6 in the Data Booklet.**

**It shows an article.**

**Caroline runs a hairdressing salon in London.**

**She reads the article in the Data Booklet.**

**Caroline states that the random variable  $X \sim B(100\,000, 0.001)$  could be used to model the number of hairs lost by a randomly selected customer during one day.**

**(continued on the next page)**

**6. continued.**

**(a) Explain how Caroline chose her parameters for the distribution of  $X$**

**(2 marks)**

**(b) Comment on the assumptions underlying Caroline's model indicating whether or not these are reasonable.**

**Your answer should include  
THREE different statements.  
(3 marks)**

**(Total for Question 6 is 5 marks)**

---

**Turn over**

**7. Refer to Table 1 and Table 2 for Question 7 in the Data Booklet.**

**Animal Crossing is a video game in which players visit desert islands.**

**A single animal lives on each desert island.**

**A player visiting an island will always meet the animal that lives there.**

**Each animal is either male or female.**

**Jonny plays a simplified version of Animal Crossing in which there are only four animal species.**

**Data on the number of animals of each species is shown in Table 1 in the Data Booklet.**

**(continued on the next page)**

**Turn over**

**7. continued.**

**Jonny visits a desert island at random.**

**(a) Write down the probability that the animal he meets is a female anteater.**

**(1 mark)**

**(b) Write down the probability that the animal he meets is a male, given that it is an alligator.**

**(1 mark)**

**(continued on the next page)**

**Turn over**

**7. continued.**

**Jonny assumes that once he has visited a particular desert island, he will never visit that island again.**

- (c) Under Jonny's assumption, find the probability that when Jonny visits four islands, he meets two birds and two bears, given that all the animals he meets are either birds or bears.**
- (5 marks)**

**(continued on the next page)**

**Turn over**

**7. continued.**

**In the full version of Animal Crossing there are MORE than four species of animal.**

**Jonny's assumption in part (c) is untrue, and it is possible to meet the same animal MULTIPLE TIMES.**

**The data for alligators, anteaters, bears and birds in Table 1 in the Data Booklet are correct for both versions of the game.**

**(continued on the next page)**

**Turn over**



**7. continued.**

**(d) Explain WHY, in the full version of the game,**

**(i) the probability in part (a) would be smaller than the calculated value**

**(1 mark)**

**(ii) the probability in part (b) would be unchanged**

**(1 mark)**

**(continued on the next page)**

**Turn over**

**7. (d) continued.**

**(iii) the probability in part (c)  
would be smaller than the  
calculated value.**

**(1 mark)**

**(continued on the next page)**

**7. continued.**

**In Animal Crossing players are able to buy and sell turnips each week.**

**The prices at which players can sell turnips during a week follow one of three patterns:**

**Spike**

**Fluctuating**

**Decreasing**

**The probability of the prices following a certain pattern one week depends only on the previous week's pattern, as shown in Table 2 in the Data Booklet.**

**(continued on the next page)**

**Turn over**

**7. continued.**

**During the first week of May, the prices of Jonny's turnips are following the fluctuating pattern.**

**(e) Find the probability that, during the THIRD week of May, the prices of Jonny's turnips are following the fluctuating pattern again.**

**(2 marks)**

**(continued on the next page)**

**7. continued.**

**During the first week of May, the prices of Jonny's turnips were following the fluctuating pattern.**

**During the third week of May, the prices of Jonny's turnips were following the fluctuating pattern again.**

**(f) Find the probability that during the SECOND week of May, the prices of Jonny's turnips were following the spike pattern.**

**(4 marks)**

**(Total for Question 7 is 16 marks)**

---

**TOTAL FOR PAPER IS 80 MARKS**

**END OF PAPER**

---