AS Statistics Practice Paper D
45 Marks

45 Minutes

1. An estate agent recorded the price per square metre, p £/m ^{2}, for 7 two-bedroom houses.

He then coded the data using the coding $q = \frac{p-a}{b}$, where a and b are positive constants.

His results are shown in the table below.

p	1840	1848	1830	1824	1819	1834	1850
q	4.0	4.8	3.0	2.4	1.9	3.4	5.0

a. Find the value of a and the value of b. **(2)**

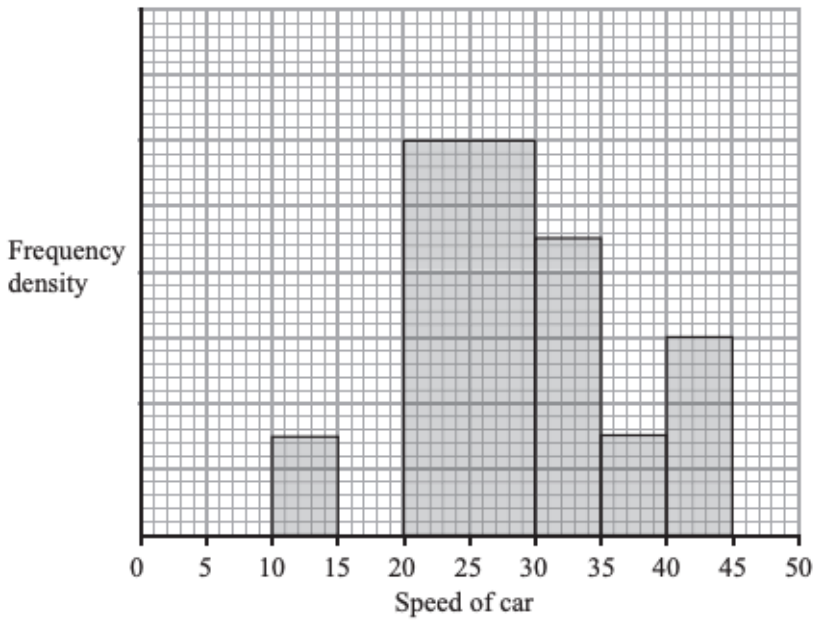
An estate agent recorded the distance, d km, of 7 two-bedroom houses from the nearest train station. The results are summarised below.

$$S_{dd} = 1.02 \qquad S_{qq} = 8.22 \qquad S_{dq} = -2.17$$

b. Calculate the mean distance of the houses **(2)**

(Total marks 4)

2.



.

A policeman records the speed of the traffic on a busy road with a 30 mph speed limit.
He records the speeds of a sample of 450 cars. The histogram in Figure 2 represents the results.

a. Calculate the number of cars that were exceeding the speed limit by at least 5 mph in the sample. **(4)**

b. Estimate the value of the mean speed of the cars in the sample. **(3)**

c. Estimate, to 1 decimal place, the value of the median speed of the cars in the sample. **(2)**

(Total marks 9)

3. In a large restaurant an average of 3 out of every 5 customers ask for water with their meal.

A random sample of 10 customers is selected.

a. Find the probability that exactly 6 ask for water with their meal. **(3)**

b. Find the probability that less than 9 ask for water with their meal. **(2)**

c. Find the smallest value of n such that

$$P(X < n) \geq 0.9$$

where the random variable X represents the number of these customers who ask for water. **(3)**

(Total marks 8)

4. A test statistic has a distribution $B(25, p)$.

a. Given that,

$$H_0: p = 0.5, \quad H_1: p \neq 0.5,$$

find the critical region for the test statistic such that the probability in each tail is as close as possible to 2.5%.

(3)

b. State the probability of incorrectly rejecting H_0 using this critical region. **(2)**

(Total marks 5)

5. A manufacturer carried out a survey of the defects in their soft toys. It is found that the probability of a toy having poor stitching is 0.03 and that a toy with poor stitching has a probability of 0.7 of splitting open. A toy without poor stitching has a probability of 0.02 of splitting open.

a. Draw a tree diagram to represent this information. **(3)**

b. Find the probability that a randomly chosen soft toy has exactly one of the two defects, poor stitching or splitting open. **(3)**

The manufacturer also finds that soft toys can become faded with probability 0.05 and that this defect is independent of poor stitching or splitting open. A soft toy is chosen at random.

c. Find the probability that the soft toy has none of these 3 defects. **(2)**

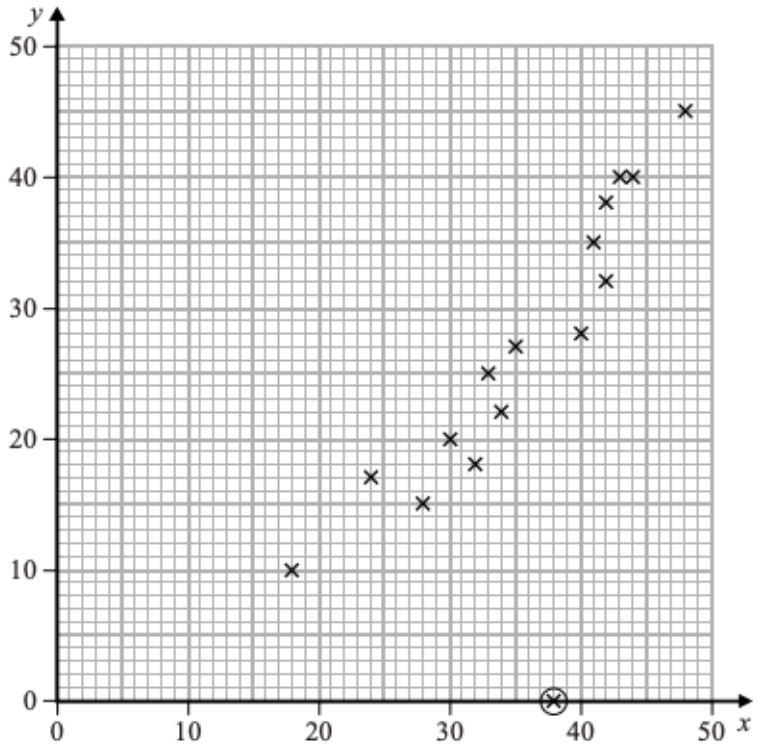
d. Find the probability that the soft toy has exactly one of these 3 defects. **(4)**

(Total marks 12)

6. Following some school examinations, Chetna is studying the results of the 16 students in her class. The mark for paper 1, x, and the mark for paper 2, y, for each student are summarised in the following statistics.

$overline\ x = 35.75$ $overline\ y = 25.75$ $\sigma_x = 7.79$ $\sigma_y = 11.91$ $sum\ xy = 15837$

Chetna decides to examine these data in more detail and plots the marks for each of the 16 students on the scatter diagram.



Chetna decides to omit the data point $(38,0)$ and examine the other 15 students' marks.

The equation of the line of regression of y on x for these 15 students is $y = 1.21x - 15.66$

Estimate the mark in the second paper for a student who scored 38 marks in the first paper. **(4)**

(Total marks 4)

7. Using the Edexcel large data set, Tom wants to take a simple random sample of size 10 for the daily maximum gust in Leuchars in October.

He uses the first two digits of the date as the sampling frame, and generates 10 random numbers between 1 and 31.

Using your knowledge of the large data set, explain why Tom might not get a sample of size 10 using his method.

(Total marks 1)

8. A factory produces components. Each component has a unique identity number and it is assumed that 2% of the components are faulty. On a particular day, a quality control manager wishes to take a random sample of 50 components.

The statistic F represents the number of faulty components in the random sample of size 50.

Specify the sampling distribution of F. **(2)**

(Total marks 2)

Total Marks for Paper: 45

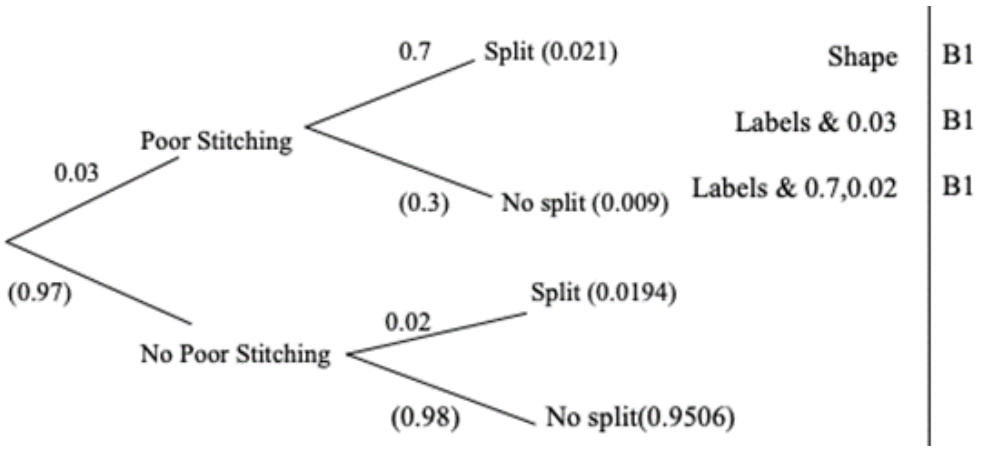
Mark Scheme

1a	$\frac{1840-a}{b} = 4.0$ $\frac{1848-a}{b} = 4.8$	**M1**
	$a = 1800$ $b = 10$	**A1**
1b	Mean = $\frac{S_{dq}}{n}$ $mean = -\frac{2.17}{7}$ $mean = -0.31$	**M1**
	The mean distance is 0.31 km from the train station.	**A1**

2a	One large square = $\frac{450}{22.5}$	**M1**
	One large square = 20 cars	**A1**
	No. > 35 is 4.5×20	**M1**
	= 90	**A1**
2b	Mean = $\frac{12975}{450}$	**M1** **M1**
	Mean = 28.83	**A1**
2c	$Q_2 = 20 + \frac{195}{240} \times 10$	**M1**
	Lower quartile = 28.125	**A1**

3a	$X \sim B(10, 0.6)$	**B1**
	$P(X = 6) = (0.6)^6(0.4)^4 \frac{10!}{6!4!}$	**M1**
	= 0.2508… =0.251	**A1**
3b	$X \sim B(10, 0.6)$	**M1**
	$P(X < 9) = 1 - P(X = 10) + P(X = 9))$ $= 1 - (0.6)^{10} - (0.6)^9(0.4)^t \frac{10!}{9!1!}$	**M1**
	= 0.9536… = 0.954	**A1**
3c	$X \sim B(50, 0.6)$	**M1**
	$Y \sim B(50, 0.4)$ $P(X < n) \geq 0.9$ $P(Y > 50 - n) \geq 0.9$ $(Y \leq 50 - n) \leq 0.1$	**M1**
	$50 - n \leq 15$ $n \geq 35$ $n = 35$	**A1**

4a	$X \sim B(25, 0.5)$	**M1**
	$P(X \leq 7) = 0.0216$ $P(X \geq 18) = 0.0216$ CR $X \leq 7; X \geq 18$	**A1** **A1**
4b	$P(rejecting\ H_0) = 0.0216 + 0.0216$	**M1**
	= 0.0432	**A1**

5a		
5b	$P(Exactly\ one\ defect) = 0.03 \times 0.3 + 0.97 \times 0.02$ $= 0.009 + 0.0194$	**M1** **A1**
	$= 0.0284$	**A1**
5c	$P(No\ defects) = (1 - 0.03) \times (1 - 0.02) \times (1 - 0.05)$	**M1**
	$= 0.90307$	**A1**
5d	$P(Exactly\ one\ defect) = (b) \times (1 - 0.05) + (1 - 0.03) \times (1 - 0.02) \times 0.05$	**M1** **M1**
	$= 0.0284 \times 0.95 + 0.97 \times 0.98 \times 0.05$	**A1**
	$= [0.02698 + 0.04753] = 0.07451$	**A1**

6	$y = 1.21\ \times 38 - 15.66$	**A1** **A1** **A1**
	= 30 marks	**A1**

7	missing OR n/a OR no data OR not applicable OR data is missing OR not available OR missing values OR data missing OR no values OR don't have values OR do not have values OR incomplete data OR values are missing OR incomplete OR might not be data OR do not have data	**M1**

8	Binomial	**B1**
	$F \sim B(50, 0.02)$	**B1**