2.10 Statistics: Reading and Interpreting Given Diagrams (Venn, Tree, Histogram, Scatter, Box Plot): 1-hour Pack

EDEXCEL 8MA0 · Calculator allowed · about 60 minutes

Total Marks

Name: _____ Date: ____ / ____ / _____

Answer ALL questions. Show all your working.

1 A histogram is drawn for the lengths of time, in minutes, that customers wait. The class widths and frequencies are:
Class: 0 - 4, 4 - 8, 8 - 12, 12 - 17
Frequency: 22, 14, 11, 27

(a) Find the frequency density for each class. **(2)**

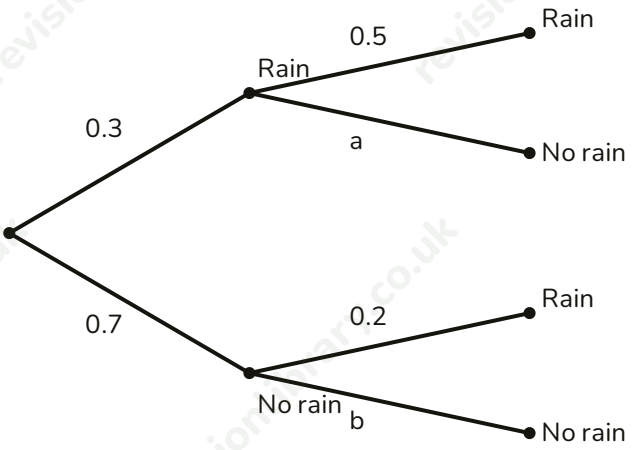
(b) Estimate the number of customers who waited longer than 6 minutes but not more than 8 minutes. **(2)**

(c) The height of the tallest bar is 6 cm. Find the height of the bar for the class with the smallest frequency density, in cm to 2 significant figures. **(2)**

(Total for Question 1 is 6 marks)

© Revision Library | revisionlibrary.co.uk

2 The tree diagram below models whether it rains on each of two consecutive days. Two of the second-stage probabilities are not shown, and are marked a and b.



(a) Write down the values of a and b. **(2)**

(b) Calculate the probability that it rains on exactly one of the two days. **(3)**

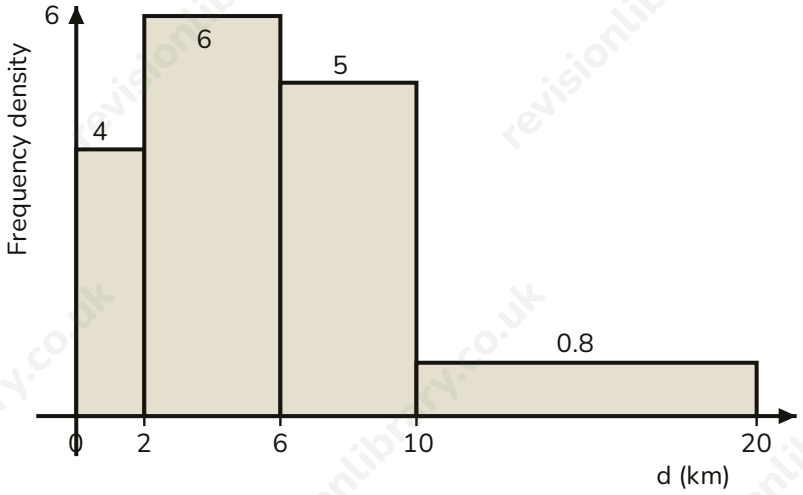
(c) Calculate the probability that it rains on at least one of the two days. **(3)**

(d) Given that it rains on at least one of the two days, find the probability that it rains on both days. **(3)**

(Total for Question 2 is 11 marks)

© Revision Library | revisionlibrary.co.uk

3 The histogram below shows the distances, d km, travelled to school by 60 students at a school.



(a) Use the histogram to find the frequency for each of the four class intervals. **(2)**

(b) Using linear interpolation, estimate the median distance travelled to school. **(3)**

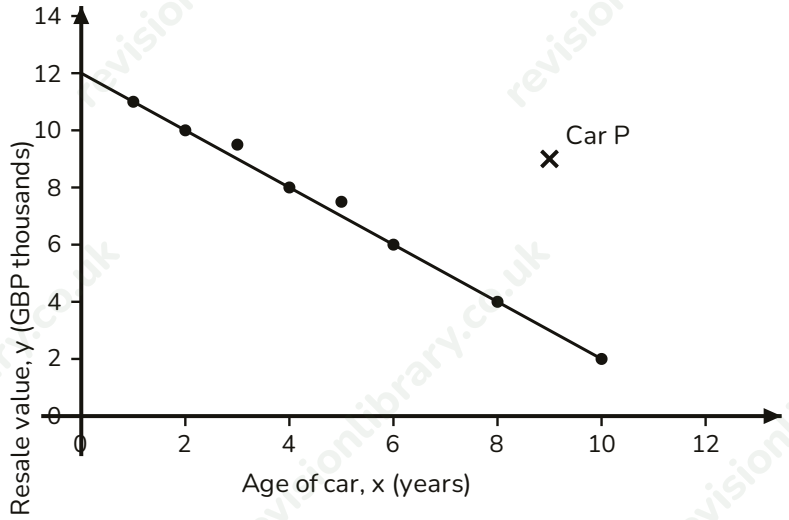
(c) Assuming distances are uniformly distributed within the class $10 \leq d < 20$, estimate the number of students who travelled more than 15 km to school. **(3)**

(d) Explain why your answer to part (c) is only an estimate. **(1)**

(Total for Question 3 is 9 marks)

© Revision Library | revisionlibrary.co.uk

4 The scatter diagram below shows the age, in years, and the resale value, in thousands of pounds, of 9 cars of the same model, together with a line of best fit.



(a) Describe the type of correlation shown between the age of a car and its resale value. **(1)**

(b) Using the line of best fit, estimate the resale value of a car aged 7 years. **(2)**

(c) Car P, aged 9 years, has a resale value of £9000, well above the value suggested by the line of best fit. Suggest a reason, in context, why this car's resale value might not fit the general trend. **(1)**

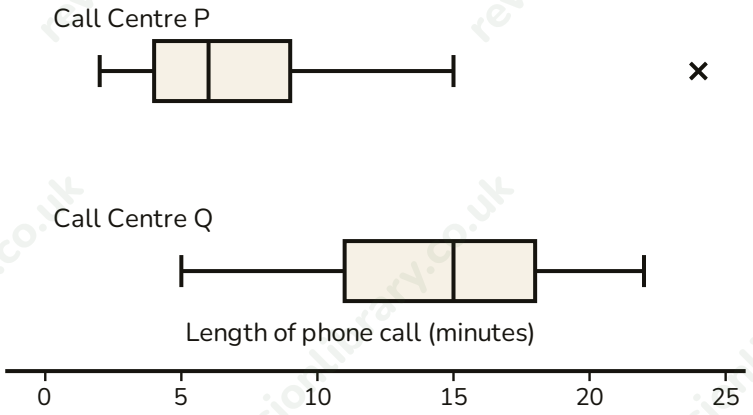
(d) State whether it would be appropriate to use the line of best fit to estimate the resale value of a car aged 20 years, giving a reason. **(2)**

(e) The line of best fit has a gradient of -1. Interpret this value in the context of the data. **(2)**

(Total for Question 4 is 8 marks)

© Revision Library | revisionlibrary.co.uk

5 The box plots below show the length of telephone calls, in minutes, handled by two call centres, P and Q, drawn on the same scale.



(a) By calculating the relevant boundary, determine whether the point at 24 minutes for Call Centre P has been correctly identified as an outlier, using the rule that a value is an outlier if it lies more than 1.5 x IQR beyond the nearer quartile. **(3)**

(b) State, giving a reason based on the position of the median within the box, the skewness of the distribution of call lengths for Call Centre P (excluding the outlier). **(2)**

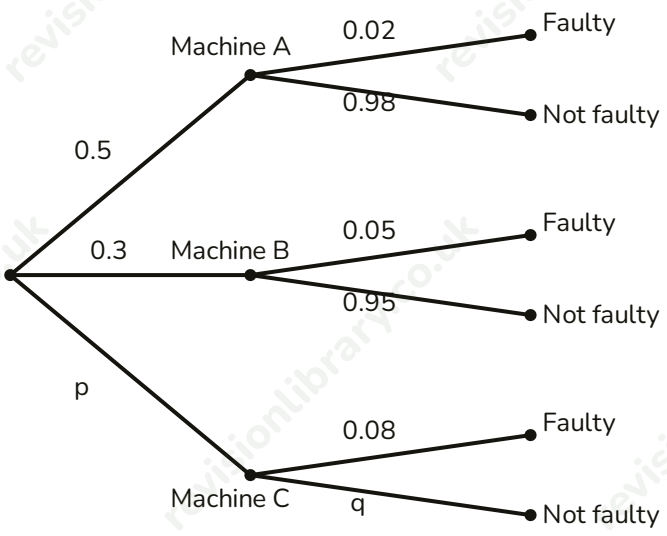
(c) State, giving a reason, the skewness of the distribution of call lengths for Call Centre Q. **(2)**

(d) A manager claims 'Call Centre Q handles calls more efficiently, since its calls are shorter on average.' Using the box plots, comment on whether the evidence supports this claim. **(3)**

(Total for Question 5 is 10 marks)

© Revision Library | revisionlibrary.co.uk

6 A factory has three machines, A, B and C, that produce identical parts. The tree diagram below shows the proportion of parts produced by each machine, and the probability that a part from each machine is faulty. Two of the probabilities are not shown, and are marked p and q.



(a) Find the values of p and q. **(2)**

(b) Calculate the probability that a randomly selected part is faulty. **(3)**

(c) Given that a randomly selected part is faulty, find the probability that it was produced by Machine C. **(3)**

(d) The factory manager claims 'Since Machine C has the highest individual fault rate (8%), most faulty parts come from Machine C.' Using your answers to parts (b) and (c), comment on whether this claim is justified. **(3)**

(Total for Question 6 is 11 marks)

© Revision Library | revisionlibrary.co.uk

Mark scheme · 2.10 Statistics: Reading and Interpreting Given Diagrams (Venn, Tree, Histogram, Scatter, Box Plot): 1-hour Pack

Question 1

(a) frequency density = frequency / class width **M1**

(a) values 11/2, 7/2, 11/4, 27/5 **A1**

(a) Answer: 11/2, 7/2, 11/4, 27/5

(b) proportion of the class 4 - 8 lying in the required interval = (8 - 6)/4 **M1**

(b) 7 **A1**

(b) Answer: 7

(c) height proportional to frequency density: scale = 6/largest fd **M1**

(c) 3.0 cm **A1**

(c) Answer: 3.0 cm

Question 2

(a) a = 0.5 **B1**

(a) b = 0.8 **B1**

(a) Answer: a = 0.5, b = 0.8

(b) identifies the two relevant outcomes: rain-then-no-rain and no-rain-then-rain **M1**

(b) correct method: 0.3 x 0.5 + 0.7 x 0.2 **M1**

(b) cao 0.29 **A1**

(b) Answer: 0.29

(c) attempts P(no rain on either day) = 0.7 x 0.8 **M1**

(c) correct method: 1 - their P(no rain on either day) **M1**

(c) cao 0.44 **A1**

(c) Answer: 0.44

(d) P(rain on both days) = 0.3 x 0.5 = 0.15 identified **M1**

(d) correct conditional method: their 0.15 divided by their answer to part (c) **M1**

(d) oe 15/44, awrt 0.341 **A1**

(d) Answer: 15/44 (awrt 0.341)

Question 3

(a) attempts frequency = frequency density x class width for at least one class **M1**

(a) all four correct: 8, 24, 20, 8 **A1**

(a) Answer: 0≤d<2: 8; 2≤d<6: 24; 6≤d<10: 20; 10≤d<20: 8

(b) correct cumulative frequencies identified (8, 32, 52, 60) and median class 2≤d<6 identified **M1**

(b) correct interpolation: 2 + ((30-8)/24) x 4 **M1**

(b) awrt 5.67 **A1**

(b) Answer: awrt 5.67 km

(c) identifies that only the proportion of the class 10≤d<20 above d=15 is needed **B1**

(c) correct method: (20-15)/(20-10) x 8 **M1**

(c) cao 4 **A1**

(c) Answer: 4 students

(d) the actual distribution of distances within the class 10≤d<20 is not known; only the total frequency for the whole class is given, so a uniform spread has to be assumed, oe **B1**

(d) Answer: Because the exact distances within the class 10≤d<20 are unknown, only the total frequency for the class; a uniform spread within the class has to be assumed.

Question 4

(a) negative correlation **B1**

(a) Answer: Negative correlation.

(b) reads the line of best fit at x = 7 appropriately **M1**

(b) awrt 5 (allow 4.5 to 5.5) **A1**

(b) Answer: awrt £5000 (£5 thousand)

(c) valid contextual reason, e.g. Car P may be a rare or collectible model, or may be in exceptionally good condition or have unusually low mileage for its age, oe **B1**

(c) Answer: Car P might be a rare or highly sought-after model, or might have very low mileage or be in exceptionally good condition for its age.

(d) no, it would not be appropriate **B1**

(d) reason: 20 years is outside the range of the given data (ages 1 to 10), so this would be extrapolation and the linear trend is not known to continue, oe **B1**

(d) Answer: No; 20 years is outside the range of the data used to draw the line (ages 1 to 10), so this would be extrapolation, and there is no evidence the linear trend continues beyond the data.

(e) states that resale value decreases by about £1000 for each extra year of age **B1**

(e) correctly links the negative sign to a decrease as age increases, oe **B1**

(e) Answer: On average, a car in this model loses about £1000 of resale value for each extra year of age.

© Revision Library | revisionlibrary.co.uk

Question 5

(a) IQR = 9 - 4 = 5 **M1**

(a) upper boundary = 9 + 1.5 x 5 **M1**

(a) cso: boundary = 16.5, and since 24 > 16.5, the point has been correctly identified as an outlier **A1**

(a) Answer: Correctly identified: the boundary is 16.5, and 24 > 16.5.

(b) compares median - Q1 (= 2) with Q3 - median (= 3) **B1**

(b) since Q3 - median > median - Q1, the distribution is positively skewed **B1**

(b) Answer: Positively skewed: Q3 - median (3) is greater than median - Q1 (2).

(c) compares median - Q1 (= 4) with Q3 - median (= 3) **B1**

(c) since median - Q1 > Q3 - median, the distribution is negatively skewed **B1**

(c) Answer: Negatively skewed: median - Q1 (4) is greater than Q3 - median (3).

(d) Call Centre Q's median (15) is higher than Call Centre P's median (6), so the typical call is longer at Q, not shorter **B1**

(d) the middle 50% of P's calls (IQR 4 to 9 minutes) are much shorter than the middle 50% of Q's calls (IQR 11 to 18 minutes), oe **B1**

(d) concludes the evidence does not support the manager's claim, oe **B1**

(d) Answer: The evidence does not support the claim: Call Centre Q's median call (15 minutes) is much higher than Call Centre P's (6 minutes), and P's middle 50% of calls (4 to 9 minutes) are far shorter than Q's (11 to 18 minutes). Overall P's calls, aside from the single outlier, tend to be shorter than Q's.

Question 6

(a) p = 0.2 **B1**

(a) q = 0.92 **B1**

(a) Answer: p = 0.2, q = 0.92

(b) identifies the need to sum P(machine) x P(faulty | machine) across all three machines **M1**

(b) correct method: 0.5 x 0.02 + 0.3 x 0.05 + 0.2 x 0.08 **M1**

(b) cao 0.041 **A1**

(b) Answer: 0.041

(c) P(C and faulty) = 0.2 x 0.08 = 0.016 identified **M1**

(c) correct conditional method: their 0.016 divided by their answer to part (b) **M1**

(c) oe 16/41, awrt 0.390 **A1**

(c) Answer: 16/41 (awrt 0.390)

(d) from part (c), Machine C accounts for about 39.0% of faulty parts, the largest single share of the three machines, but this is not a majority **B1**

(d) notes that Machine B contributes a comparable share (0.015/0.041, about 36.6%), so Machine C is not overwhelmingly dominant **B1**

(d) balanced conclusion: the claim that 'most' faulty parts come from Machine C is not justified, since this is under half of all faulty parts, oe **B1**

(d) Answer: Not justified: Machine C does contribute the largest single share of faulty parts (about 39.0%, from part (c)), but this is well under half. Machine B contributes a similar share (about 36.6%), so the claim that most faulty parts come from Machine C is incorrect.

© Revision Library | revisionlibrary.co.uk