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

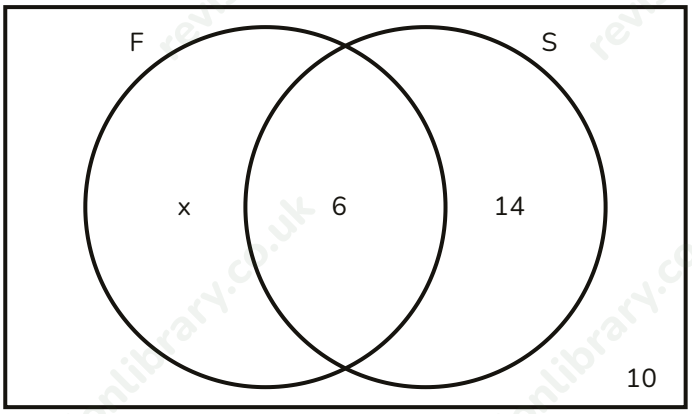
EDEXCEL 8MA0 · Calculator allowed · about 70 minutes

Total Marks

Name: _____ Date: ____ / ____ / _____

Answer ALL questions. Show all your working.

1 A Venn diagram is used to show data about 50 sixth-form students, some of whom study French (F) and some of whom study Spanish (S). The number of students in each region of the diagram is shown, except for one region, which is marked x.



(a) Using the fact that the diagram represents 50 students in total, find the value of x. **(2)**

(b) Find P(a randomly selected student studies both French and Spanish), giving your answer as a fraction in its simplest form. **(2)**

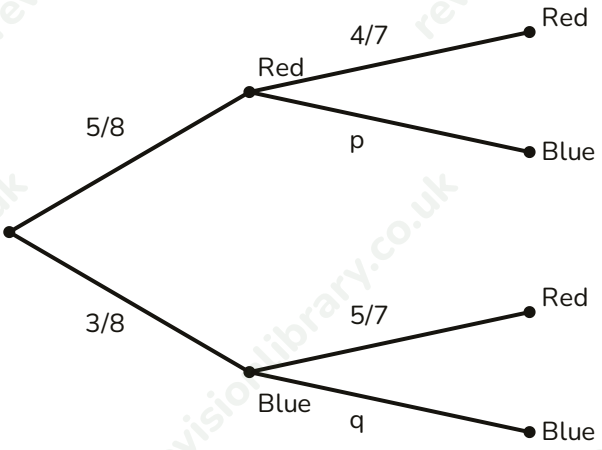
(c) Find P(a student studies French, given that they study Spanish). **(3)**

(d) Determine, showing your working, whether the events 'a student studies French' and 'a student studies Spanish' are independent. **(3)**

(Total for Question 1 is 10 marks)

© Revision Library | revisionlibrary.co.uk

2 A bag contains 5 red counters and 3 blue counters. Two counters are drawn from the bag at random, one after the other, without replacement. The tree diagram below shows the possible outcomes. Two of the second-stage probabilities are not shown, and are marked p and q.



(a) Write down the value of p and the value of q. **(2)**

(b) Calculate P(both counters are red). **(2)**

(c) Calculate the probability that the two counters are different colours. **(3)**

(d) Given that the two counters drawn are the same colour, find the probability that both are red. **(4)**

(Total for Question 2 is 11 marks)

© Revision Library | revisionlibrary.co.uk

3 The masses, in kg, of a sample of lambs have lower quartile 21, median 26, upper quartile 30, smallest value 15 and largest value 62.

(a) Find the interquartile range. **(2)**

(b) A value is an outlier if it is more than 1.5 x IQR above the upper quartile or below the lower quartile. Determine whether the smallest value and the largest value are outliers. **(3)**

(c) Describe the skewness of the distribution, justifying your answer using the quartiles. **(2)**

(Total for Question 3 is 7 marks)

---

4 A tree diagram models whether a patient is in group A (probability 2/3) or group B, and whether the patient recovers (R). The probability of recovering is 2/3 for group A and 1/3 for group B.

(a) Write down the probability that a patient is in group B, and the probability that a group B patient does not recover. **(2)**

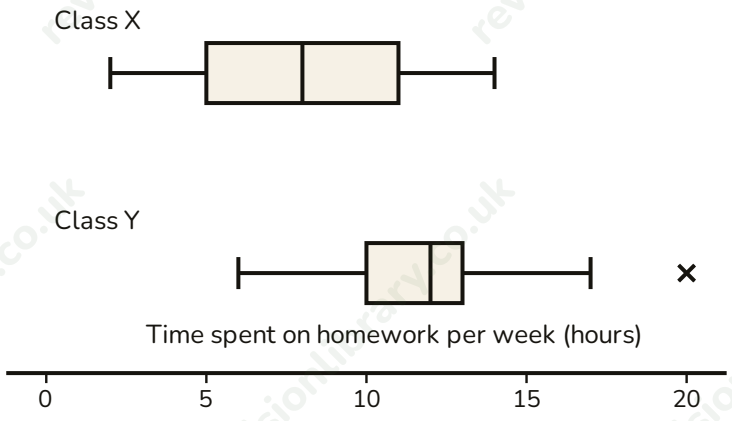
(b) Find the probability that a randomly chosen patient recovers. **(2)**

(c) Given that a patient recovered, find the probability that the patient was in group A. **(2)**

(Total for Question 4 is 6 marks)

© Revision Library | revisionlibrary.co.uk

5 The box plots below show the time spent on homework per week, in hours, by two classes of sixth-form students, Class X and Class Y, drawn on the same scale.



(a) Write down the median and calculate the interquartile range for Class X. **(3)**

(b) State the five-figure summary for Class Y, including the value marked as an outlier as the maximum. **(2)**

(c) Explain why the upper whisker on the box plot for Class Y is drawn to 17 rather than extending to 20. **(2)**

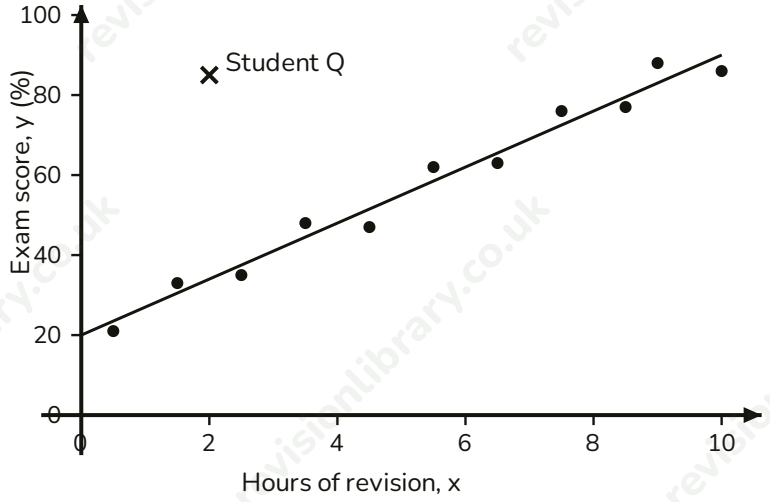
(d) Using your answers to parts (a) and (b), compare the median times and the interquartile ranges of the two classes, in context. **(3)**

(e) State one advantage of using box plots, rather than the mean and standard deviation, to compare the two classes' homework times. **(1)**

(Total for Question 5 is 11 marks)

© Revision Library | revisionlibrary.co.uk

6 The scatter diagram below shows the number of hours of revision carried out by 12 students before an exam, and the percentage score each student achieved, together with a line of best fit. The product moment correlation coefficient (PMCC) for the 12 students is also shown.



r = 0.78

(a) State, with a reason, whether the value r = 0.78 shown on the diagram indicates a strong or a weak positive correlation. **(2)**

(b) Using the line of best fit, estimate the exam score of a student who revised for 6 hours. **(2)**

(c) Student Q revised for only 2 hours but scored 85%, well above the value suggested by the line. State the effect that removing Student Q's data point would be likely to have on the value of r, and explain your reasoning. **(2)**

(d) A teacher claims 'revising for longer causes a higher exam score.' Using the scatter diagram, comment on whether this conclusion is fully justified. **(3)**

(Total for Question 6 is 9 marks)

© Revision Library | revisionlibrary.co.uk

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

Question 1

(a) correct method: x = 50 - 6 - 14 - 10 **M1**

(a) cao x = 20 **A1**

(a) Answer: x = 20

(b) 6/50 identified from the diagram **M1**

(b) cao 3/25 **A1**

(b) Answer: 3/25

(c) n(S) = 6 + 14 = 20 identified from the diagram **M1**

(c) correct method: 6/20 **M1**

(c) cao 3/10 **A1**

(c) Answer: 3/10

(d) P(F) = (20 + 6)/50 = 26/50 (= 13/25 or 0.52) found **M1**

(d) P(F) x P(S) = 0.52 x 0.4 = 0.208 attempted, using P(S) = 20/50 = 0.4 **M1**

(d) cso: compares 0.208 with P(F and S) = 6/50 = 0.12, correctly concludes the values are not equal so F and S are not independent **A1**

(d) Answer: Not independent, since P(F) x P(S) = 0.208 does not equal P(F and S) = 0.12.

Question 2

(a) p = 3/7 **B1**

(a) q = 2/7 **B1**

(a) Answer: p = 3/7, q = 2/7

(b) correct method: 5/8 x 4/7 **M1**

(b) oe 5/14 **A1**

(b) Answer: 5/14

(c) identifies the two relevant outcomes: red-then-blue and blue-then-red **M1**

(c) correct method: 5/8 x 3/7 + 3/8 x 5/7 **M1**

(c) cao 15/28 **A1**

(c) Answer: 15/28

(d) attempt P(both red) = 5/8 x 4/7 (= 5/14 or 20/56) **M1**

(d) attempt P(both blue) = 3/8 x 2/7 (= 3/28 or 6/56) **M1**

(d) correct conditional method: P(both red) / [P(both red) + P(both blue)], dependent on both previous marks **dM1**

(d) cao 10/13 **A1**

(d) Answer: 10/13

Question 3

(a) 30 - 21 **M1**

(a) 9 **A1**

(a) Answer: 9

(b) lower limit 21 - 1.5 x 9 = 7.5 and upper limit 30 + 1.5 x 9 = 43.5 **M1**

(b) smallest value 15 is not below 7.5, so is not an outlier **A1**

(b) largest value 62 is above 43.5, so is an outlier **A1**

(b) Answer: smallest is not an outlier; largest is an outlier

(c) compares median - Q1 = 5 and Q3 - median = 4 **M1**

(c) negative skew (as the larger spread is on that side) **A1**

(c) Answer: negative skew

Question 4

(a) P(B) = 1/3 **B1**

(a) P(not R | B) = 2/3 **B1**

(a) Answer: P(B) = 1/3, P(not R | B) = 2/3

(b) P(A and R) + P(B and R) **M1**

(b) 2/3 x 2/3 + 1/3 x 1/3 = 5/9 **A1**

(b) Answer: 5/9

(c) P(A | R) = P(A and R)/P(R) **M1**

(c) 4/5 **A1**

(c) Answer: 4/5

Question 5

(a) median = 8 **B1**

(a) correct method: 11 - 5 **M1**

(a) cao IQR = 6 **A1**

(a) Answer: median = 8 hours, IQR = 6 hours

(b) minimum = 6, lower quartile = 10, median = 12, upper quartile = 13, all correct **B1**

(b) maximum correctly identified as 20 (i.e. including the outlier as the true maximum value) **B1**

(b) Answer: minimum = 6, Q1 = 10, median = 12, Q3 = 13, maximum = 20

(c) recognises that 20 is identified as an outlier (it lies more than 1.5 x IQR beyond the upper quartile) **B1**

(c) states that the whisker is drawn only as far as the largest value that is not classified as an outlier, which is 17, oe **B1**

(c) Answer: Because 20 is classified as an outlier, it is plotted separately rather than as the end of the whisker; the whisker extends only to the largest value that is not an outlier, 17.

(d) Class Y has the higher median (12 compared with 8), so Class Y students typically spent more time on homework than Class X, oe **B1**

(d) compares Class X's IQR (their value from part (a), = 6) with Class Y's IQR (13 - 10 = 3): Class X's IQR is the larger of the two, so homework times were more variable in Class X than in Class Y, oe **B1ft**

(d) valid overall concluding comment in context, e.g. Class Y students spent more time on homework on average and were also more consistent in how long they spent than Class X, oe **B1**

(d) Answer: Class Y has the higher median (12 vs 8), so typically spent more time on homework; Class X has the larger IQR (6 vs 3), so homework times were more spread out in Class X. Overall, Class Y students did more homework and more consistently than Class X.

(e) box plots directly show the median, quartiles and any outliers, and are not distorted by extreme values (outliers) the way the mean and standard deviation can be, oe **B1**

(e) Answer: Box plots show the median, quartiles and any outliers directly, and are not distorted by extreme values the way the mean and standard deviation can be.

© Revision Library | revisionlibrary.co.uk

Question 6

(a) strong (positive correlation) **B1**

(a) reason: r = 0.78 is close to 1, oe **B1**

(a) Answer: Strong positive correlation, since r = 0.78 is close to 1.

(b) reads the line of best fit at x = 6 appropriately **M1**

(b) awrt 62 (allow 60 to 64) **A1**

(b) Answer: awrt 62%

(c) r would increase (the correlation would become stronger) **B1**

(c) reason: Student Q's point does not follow the general upward trend and is currently weakening the correlation, oe **B1**

(c) Answer: r would increase, since Student Q's point does not fit the general trend and is currently pulling the correlation coefficient down.

(d) correlation does not necessarily imply causation **B1**

(d) identifies a plausible confounding variable, e.g. a student's general ability or motivation, which could affect both revision time and exam score **B1**

(d) balanced conclusion: the teacher's claim is not fully justified by the correlation alone, oe **B1**

(d) Answer: A strong correlation does not prove causation. A student's ability or motivation could independently cause both more revision and a higher score, so the teacher's causal claim is not fully justified by this data alone.

© Revision Library | revisionlibrary.co.uk