# 2.6 Statistics: Data Presentation and Interpretation Depth: Test Pack

EDEXCEL 9MA0 · Calculator allowed · about 60 minutes

Total Marks

Name: _____ Date: ____ / ____ / _____

**Answer ALL questions. Show all your working.**

## A-Level Statistics: Data Presentation and Interpretation - Depth Practice

*Original material written for Revision Library, styled on Edexcel AS/A-level Mathematics (9MA0) Statistics content.*

This pack extends core data presentation and interpretation skills to full A-level depth: measures of location and spread (including coding), identifying outliers by the quartile rule and the standard deviation rule, box plots and skewness, histograms and frequency density, linear interpolation, correlation and regression (including coding and residuals), and evaluating statistical models in context, in the style of the large data set used across the qualification. All contexts and figures are original.

**1** A tutor records the daily rainfall, in mm, at a weather station near Leeds over 8 consecutive days in March:

12, 15, 9, 21, 18, 14, 11, 16

**(a)** Calculate the mean and the standard deviation of the rainfall for these 8 days, giving each answer to 3 significant figures where appropriate. **(4)**

**(b)** The 8 rainfall values are converted from millimetres to centimetres. State, without further calculation, the new mean and the new standard deviation. **(2)**

**(Total for Question 1 is 6 marks)**

---

**2** A scientist records data on temperature, t degrees Celsius, and the number of ice creams sold, n, for several days. The regression line of n on t is $n = 50 + 2.5t$. The product moment correlation coefficient is $r = 0.85$.

**(a)** Interpret the value of the gradient of the regression line in context. **(1)**

**(b)** Use the regression line to estimate n when $t = 27$, and comment on the reliability of this estimate. **(3)**

**(c)** Comment on the strength and direction of the linear correlation indicated by $r = 0.85$. **(2)**

**(Total for Question 2 is 6 marks)**


© Revision Library | revisionlibrary.co.uk

**3** A histogram is drawn for the lengths of time, in minutes, that customers wait. The class widths and frequencies are:

Class: 0 - 2, 2 - 7, 7 - 12, 12 - 14

Frequency: 5, 5, 10, 22

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

**(b)** Estimate the number of customers who waited longer than 4.5 minutes but not more than 7 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 3 is 6 marks)**

---

**4** The times, in minutes, taken by students to complete the same homework task were summarised for two schools using the following five-figure summaries.

School A: minimum = 5, Q1 = 18, median = 25, Q3 = 30, maximum = 34

School B: minimum = 9, Q1 = 15, median = 17, Q3 = 26, maximum = 40

**(a)** Calculate the interquartile range for each school. **(2)**

**(b)** By comparing the position of the median relative to the quartiles (or the lengths of the whiskers) for each school, describe and justify the skewness of each distribution. **(4)**

**(c)** Using an appropriate measure, compare the overall variation (spread) in completion times between the two schools, and comment on whether your two measures give the same conclusion. **(2)**

**(Total for Question 4 is 8 marks)**


© Revision Library | revisionlibrary.co.uk

**5** A student, Kwame, is investigating the relationship between the number of GCSE subjects studied at higher tier, x, and the total revision hours per week, y, for a sample of 10 students. To simplify the calculations, the data is coded using $u = x - 8$ and $v = y - 20$, giving the following summary statistics for the coded data:

$n = 10$, $\text{sum}(u) = 14$, $\text{sum}(v) = 35$, $\text{sum}(u^2) = 54$, $\text{sum}(v^2) = 205$, $\text{sum}(uv) = 97$

**(a)** Show that Suv, the value of Sxy for the coded data, is equal to 48. **(2)**

**(b)** Given that Suu = 34.4, find the equation of the regression line of v on u in the form $v = a + bu$, giving a and b to 3 significant figures. **(4)**

**(c)** Hence find the equation of the regression line of y on x, in the form $y = c + dx$, giving c and d to 3 significant figures. **(3)**

**(d)** Use your equation from part (c) to estimate the total revision hours per week for a student studying 10 GCSE subjects at higher tier. **(2)**

**(Total for Question 5 is 11 marks)**

---

**6** A researcher models the relationship between the number of hours of sunshine, h, and the maximum daily temperature, t degrees Celsius, recorded at a UK coastal town over a 15-day period in July. The regression line of t on h is found to be

$t = 14.2 + 0.85h$

**(a)** Interpret the value 0.85 in the context of the model. **(2)**

**(b)** Interpret the value 14.2 in the context of the model, and comment on the reliability of this interpretation. **(2)**

**(c)** Use the regression equation to estimate the maximum temperature on a day with 6 hours of sunshine. **(2)**

**(d)** Explain why it would not be appropriate to use this equation to estimate the maximum temperature on a day with 20 hours of sunshine. **(2)**

**(e)** One value of h in the data set was recorded as 0.2 hours, with an actual temperature of 20.8 degrees Celsius. Calculate the residual for this data point, and comment on what this suggests. **(3)**

**(Total for Question 6 is 11 marks)**


© Revision Library | revisionlibrary.co.uk

# Mark scheme · 2.6 Statistics: Data Presentation and Interpretation Depth: Test Pack

A correct answer earns full marks unless the question asks for working, a method, proof or explanation.

## Question 1

(a) sum of x = 116 seen, or mean attempted **M1**

(a) mean = 14.5 (mm) cao **A1**

(a) sum of $x^2$ = 1788 seen and Sxx = sum($x^2$) - (sum x)$^2$/n used (= 106) **M1**

(a) standard deviation = awrt 3.64 (mm) **A1**

**(a) Answer: Mean = 14.5 mm, standard deviation = 3.64 mm (3sf)**

(b) new mean = 1.45 (cm) ft from (a) **B1**

(b) new standard deviation = 0.364 (cm) ft from (a) **B1**

**(b) Answer: New mean = 1.45 cm, new standard deviation = 0.364 cm**

## Question 2

(a) each extra degree in t is associated with a change of +2.5 in n, so as t increases n increases **B1**

**(a) Answer: gradient 2.5: n increases by 2.5 for each unit increase in t**

(b) substitutes t = 27 **M1**

(b) n = 117.5 (accept 118 to the nearest whole number) **A1**

(b) reliable only if 27 is within the range of the data (interpolation); extrapolation is unreliable **B1**

**(b) Answer: 117.5 (118 to the nearest whole number)**

(c) strong **B1**

(c) positive correlation **B1**

**(c) Answer: strong positive linear correlation**

## Question 3

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

(a) values 5/2, 1, 2, 11 **A1**

**(a) Answer: 5/2, 1, 2, 11**

(b) proportion of the class 2 - 7 lying in the required interval = (7 - 4.5)/5 **M1**

(b) 2.5 **A1**

**(b) Answer: 2.5**

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

(c) 0.55 cm **A1**

**(c) Answer: 0.55 cm**

## Question 4

(a) IQR for School A = 12 **B1**

(a) IQR for School B = 11 **B1**

**(a) Answer: IQR(A) = 12 minutes, IQR(B) = 11 minutes**

(b) School A is negatively skewed (skewed to the left) **B1**

(b) valid justification for School A, e.g. median - Q1 = 7 > Q3 - median = 5, or the lower whisker (13) is longer than the upper whisker (4) **B1**

(b) School B is positively skewed (skewed to the right) **B1**

(b) valid justification for School B, e.g. Q3 - median = 9 > median - Q1 = 2, or the upper whisker (14) is longer than the lower whisker (6) **B1**

**(b) Answer: School A is negatively skewed; School B is positively skewed**

(c) valid comparison using both IQR (12 vs 11) and range (29 vs 31), with correct values shown **M1**

(c) correct conclusion: the IQRs are similar, suggesting comparable spread in the middle 50% of the data, but School B has the larger overall range (31 vs 29), so the two measures do not fully agree, with School B's spread more affected by its extreme values **A1**

**(c) Answer: IQR is similar for both schools (12 vs 11), but School B has the larger range (31 vs 29); the two measures do not fully agree**

## Question 5

(a) Suv = sum(uv) - [sum(u) x sum(v)]/n substituted correctly **M1**

(a) cso: Suv = 97 - (14 x 35)/10 = 97 - 49 = 48 **A1**

**(a) Answer: Suv = 48**

(b) Suu = 34.4 confirmed/used (= 54 - $14^2$/10) **B1**

(b) b = Suv/Suu = 48/34.4 attempted **M1**

(b) b = awrt 1.40 **A1**

(b) a = mean(v) - b x mean(u) = 3.5 - 1.3953(1.4) = awrt 1.55 (ft on b) **A1**

**(b) Answer: v = 1.55 + 1.40u (3sf)**

(c) substitutes u = x - 8 and v = y - 20 into v = a + bu **M1**

(c) rearranges to make y the subject: y = 20 + a + b(x-8) **M1**

(c) y = awrt 10.4 + awrt 1.40x (using unrounded a, b: y = 10.4 + 1.40x, 3sf) **A1**

**(c) Answer: y = 10.4 + 1.40x (3sf)**

(d) substitutes x = 10 into the equation from part (c), ideally using unrounded coefficients **M1**

(d) y = awrt 24.3 hours (accept 24.3-24.4 depending on rounding used, ft candidate's (c)) **A1**

**(d) Answer: Approximately 24.3 hours**


© Revision Library | revisionlibrary.co.uk

## Question 6

(a) correct rate stated: for each additional hour of sunshine **B1**

(a) the maximum daily temperature is estimated to increase by 0.85 degrees Celsius, oe **B1**

**(a) Answer: For each extra hour of sunshine, the maximum temperature is estimated to increase by 0.85 degrees Celsius**

(b) 14.2 degrees Celsius is the model's estimate of the maximum temperature when there are 0 hours of sunshine **B1**

(b) this is unreliable, since h = 0 is likely to be outside (or at the extreme edge of) the range of data collected in July at a UK coastal town, so this is an extrapolation, oe **B1**

**(b) Answer: 14.2 is the estimated temperature with 0 hours of sunshine; this is unreliable as it is an extrapolation beyond the likely range of the July data**

(c) substitutes h = 6 into t = 14.2 + 0.85h **M1**

(c) t = 19.3 degrees Celsius **A1**

**(c) Answer: 19.3 degrees Celsius**

(d) 20 hours is outside the range of the data used to form the model (extrapolation) **B1**

(d) correct contextual reasoning: a UK day in July has at most around 16-17 hours of daylight, so 20 hours of sunshine is not physically possible and the linear model has no evidence of being valid there **B1**

**(d) Answer: 20 hours of sunshine is beyond the range of the data (and beyond the hours of daylight available), so the model cannot reliably be extrapolated there**

(e) predicted t = 14.2 + 0.85(0.2) = 14.37 found **M1**

(e) residual = actual - predicted = 20.8 - 14.37 = 6.43 **M1**

(e) correct comment: the large positive residual shows this point lies well above the regression line, i.e. it is poorly fitted by the model (possibly an anomalous or unusually warm, low-sunshine day) **A1**

**(e) Answer: Residual = 6.43 degrees Celsius; the model fits this point poorly, with the actual temperature far higher than predicted**


© Revision Library | revisionlibrary.co.uk