

GCSE Statistics

Exam Insights June 2024

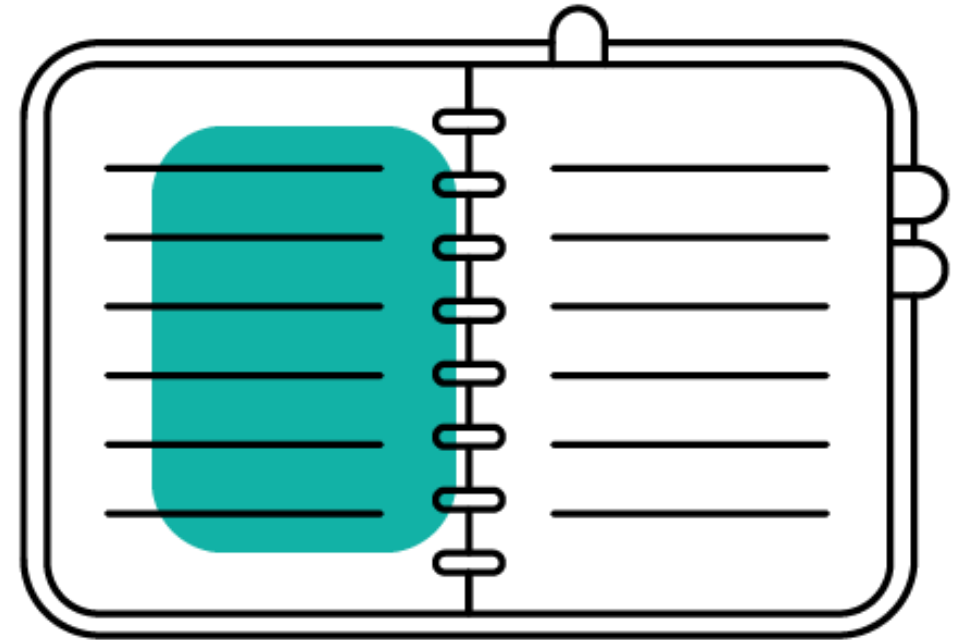
Melanie Muldowney
Christian Seager



Agenda

In this session we are going to look at:

- Entry data
- Paper 1 & Paper 2
 - Winners
 - Losers... Areas for improvement?
- Examiner key notes
- Grade boundaries
- Ongoing support – what's available now



Entry Data

From JCQ



(All UK Candidates)

	Year	Number Sat	7/A	4/C	1/G	U
Statistics	2024	31844	20.0	70.3	97.3	100
Statistics	2023	26559	20.5	71.0	97.7	100
Statistics	2022	22066	28.0	77.8	98.4	100
Statistics	2019	24027	19.3	72.9	98.0	100

82.7 % of entries (up from 81.7% in 2023)

Edexcel

1ST0		9	8	7	6	5	4	3	2	1
Foundation tier	13653					18.2	44.3	67.8	84.0	95.5
Higher tier	12686	8.3	22.9	41.4	63.6	85.0	97.0	99.2		
Total	26339	4.0	11.0	19.9	30.6	50.4	69.7	82.9	91.3	97.3

Pearson Tiers of Entry

	Summer 24	Summer 23	Summer 22	Summer 21
Foundation	13653 (51.8%)	9721 (45%)	7215 (40%)	5391 (37%)
Higher	12686	12000	10807	9358
Total	26339	21721	18022	14749

Are Year 10s affecting this? Older students ? No ... more whole cohorts?

(All UK Candidates - aged 15 and under)

Subject	Year	Number Sat	7/A	4/C	1/G	U
Statistics	2024	8429	28.0	76.2	98.2	100
Statistics	2023	8449	25.4	74.4	98.7	100
Statistics	2022	8586	29.0	78.5	98.8	100
Statistics	2019	9918	18.3	73.0	98.3	100

(All UK Candidates - aged 17 and over)

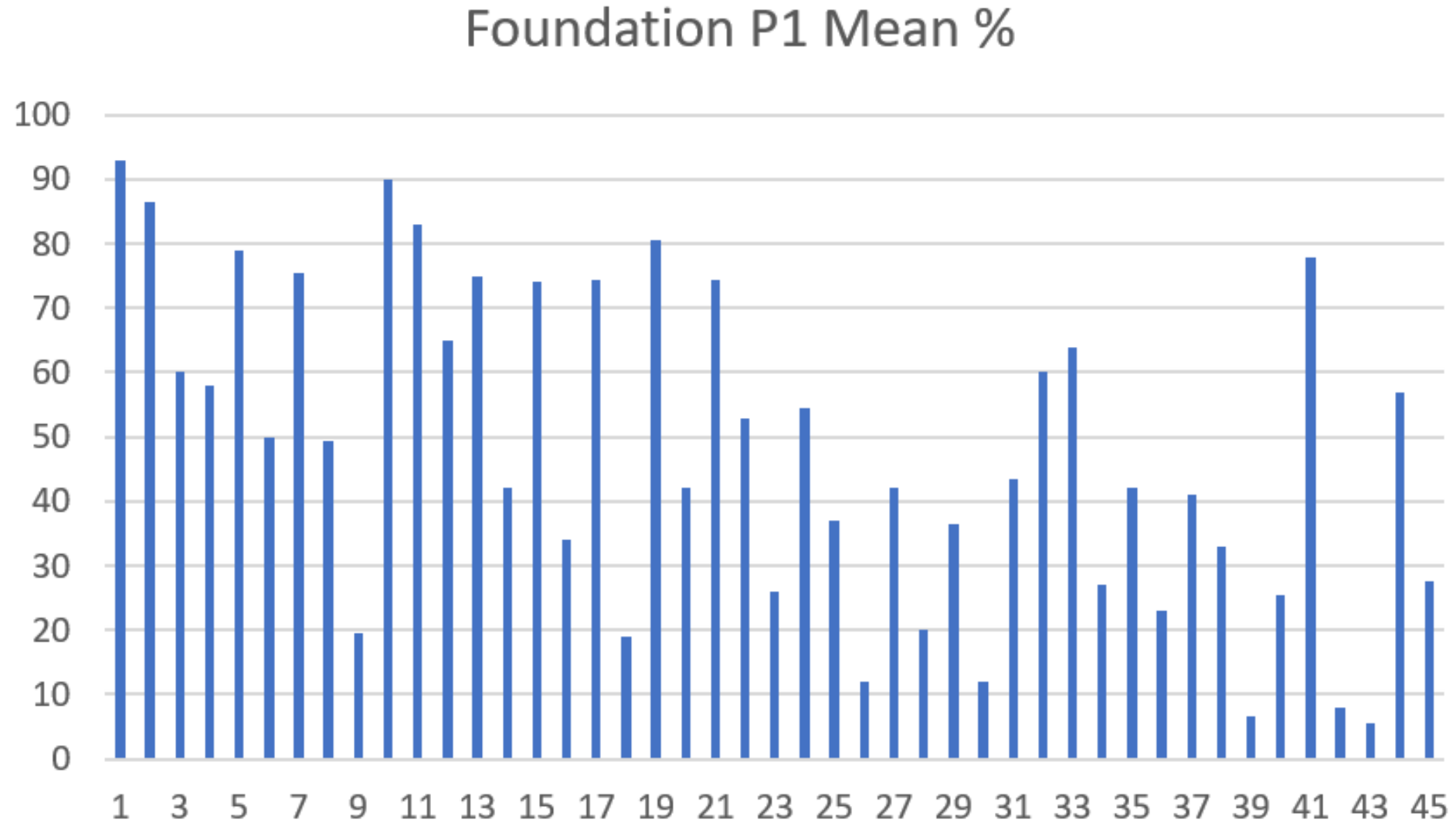
Subject	Year	Number Sat	7/A	4/C	1/G	U
Statistics	2024	520	20.2	79.2	99.6	100
Statistics	2023	500	19.4	76.2	98.0	100
Statistics	2022	388	25.8	78.1	98.5	100
Statistics	2019	381	15.7	67.5	95.5	100

Foundation Paper 1



A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
Q	Part	Position in paper	Performance	Difference	Skill tested	Mean score	Max score	Mean %	5	4	3	2	1	U	ALL
Q01	a	1	1	0	Pictogram - Completing	0.93	1	93	0.98	0.98	0.97	0.92	0.81	0.51	0.93
Q01	b	2	3	-1	Pictogram - Using	1.73	2	87	1.92	1.87	1.82	1.68	1.35	0.64	1.73
Q01	c	3	15	-12	Pictogram - Comparing	1.20	2	60	1.47	1.36	1.26	1.08	0.75	0.27	1.20
Q01	d	4	17	-13	Importance of reliability and validity	1.16	2	58	1.67	1.47	1.18	0.82	0.43	0.11	1.16
Q02	ai	5	6	-1	Bar charts - completing	0.79	1	79	0.97	0.94	0.85	0.69	0.38	0.07	0.79
Q02	aii	6	21	-15	Bar charts - completing	0.50	1	50	0.78	0.65	0.48	0.33	0.15	0.02	0.50
Q02	b	7	8	-1	Bar charts - completing	1.51	2	76	1.88	1.77	1.62	1.36	0.80	0.16	1.51
Q02	c	8	22	-14	Bar charts - Comparing	0.99	2	50	1.58	1.25	0.94	0.67	0.34	0.05	0.99
Q02	d	9	39	-30	Justify the appropriateness	0.39	2	20	0.85	0.49	0.29	0.16	0.05	0.01	0.39
Q03	ai	10	2	8	Tabulation - Reading	0.90	1	90	0.97	0.94	0.92	0.89	0.80	0.42	0.90
Q03	aii	11	4	7	Tabulation - Reading	0.83	1	83	0.93	0.90	0.86	0.80	0.65	0.26	0.83
Q03	b	12	13	-1	Tabulation - Interpreting	0.65	1	65	0.84	0.75	0.65	0.56	0.39	0.10	0.65
Q03	c	13	9	4	Tabulation - Comparing	1.50	2	75	1.77	1.70	1.58	1.39	1.01	0.28	1.50
Q03	d	14	24	-10	Time series	0.84	2	42	1.35	1.08	0.84	0.52	0.16	0.01	0.84
Q03	e	15	12	3	Identify trends in data	0.74	1	74	0.90	0.86	0.80	0.67	0.39	0.09	0.74
Q04	a	16	31	-15	Definitions of different types of data	0.34	1	34	0.53	0.39	0.31	0.24	0.20	0.10	0.34
Q04	b	17	10	7	Calculate median	1.49	2	75	1.84	1.74	1.59	1.35	0.80	0.13	1.49
Q04	c	18	40	-22	Justify the rationale for using median	0.19	1	19	0.32	0.21	0.16	0.14	0.09	0.03	0.19
Q04	d	19	5	14	Calculate range	1.61	2	81	1.97	1.93	1.80	1.40	0.73	0.09	1.61
Q04	e	20	25	-5	Compare data sets using range and median	1.68	4	42	3.17	2.32	1.52	0.70	0.14	0.01	1.68
Q05	a	21	11	10	Sample space diagrams - completing	1.49	2	75	1.92	1.80	1.57	1.24	0.70	0.22	1.49
Q05	bi	22	20	2	Sample space diagrams - using	0.53	1	53	0.89	0.73	0.51	0.28	0.09	0.01	0.53
Q05	bii	23	35	-12	Sample space diagrams - using	0.26	1	26	0.61	0.34	0.17	0.08	0.03	0.00	0.26
Q05	c	24	19	5	Experimental v theoretical values	1.09	2	55	1.70	1.40	1.11	0.71	0.24	0.04	1.09
Q06	a	25	29	-4	Hypothesis testing	0.37	1	37	0.69	0.50	0.33	0.16	0.07	0.01	0.37
Q06	b	26	41	-15	Systematic sampling	0.12	1	12	0.31	0.17	0.07	0.03	0.00	0.00	0.12
Q06	ci	27	26	1	Systematic sampling	0.42	1	42	0.69	0.56	0.39	0.23	0.09	0.01	0.42
Q06	cii	28	38	-10	Systematic sampling	0.40	2	20	1.02	0.58	0.22	0.05	0.02	0.00	0.40
Q06	d	29	30	-1	Questionnaires	0.73	2	37	1.05	0.89	0.76	0.57	0.28	0.06	0.73
Q06	ei	30	42	-12	Planning data collection	0.12	1	12	0.29	0.15	0.07	0.03	0.01	0.00	0.12
Q06	eii	31	23	8	Planning data collection	0.87	2	44	1.39	1.13	0.84	0.56	0.25	0.03	0.87
Q06	f	32	16	16	Reliability and validity	0.60	1	60	0.80	0.73	0.64	0.48	0.25	0.04	0.60
Q07	a	33	14	19	Pie chart - comparing	1.28	2	64	1.66	1.55	1.39	1.05	0.52	0.10	1.28
Q07	b	34	34	0	Pie chart	0.54	2	27	1.29	0.76	0.33	0.10	0.02	0.00	0.54
Q08	a	35	27	8	Tabulation - Reading	0.84	2	42	1.56	1.19	0.74	0.32	0.07	0.00	0.84
Q08	bi	36	37	-1	Arithmetic mean	0.69	3	23	1.73	0.91	0.43	0.16	0.04	0.00	0.69
Q08	bii	37	28	9	Arithmetic mean	0.41	1	41	0.77	0.56	0.36	0.17	0.05	0.01	0.41
Q08	biii	38	32	6	Arithmetic mean	0.33	1	33	0.68	0.46	0.27	0.12	0.04	0.01	0.33
Q08	c	39	44	-5	Select and justify appropriate diagrams	0.13	2	7	0.40	0.15	0.05	0.02	0.01	0.00	0.13
Q09		40	36	4	Population pyramid	1.27	5	25	2.65	1.68	1.05	0.48	0.11	0.00	1.27
Q10	a	41	7	34	Histograms- using	1.56	2	78	1.94	1.88	1.74	1.34	0.66	0.14	1.56
Q10	b	42	43	-1	Histograms - using	0.16	2	8	0.46	0.17	0.08	0.06	0.03	0.01	0.16
Q10	c	43	45	-2	Skewness	0.11	2	6	0.29	0.14	0.08	0.03	0.01	0.00	0.11
Q11	a	44	18	26	Tree diagrams	1.14	2	57	1.82	1.54	1.15	0.64	0.16	0.02	1.14
Q11	b	45	33	12	Formal notation for independent events	1.10	4	28	3.10	1.47	0.50	0.16	0.02	0.01	1.10
						36.53	80	46	56.40	45.04	35.29	25.44	14.19	4.08	36.53

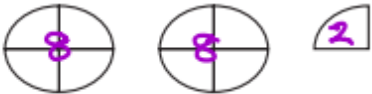
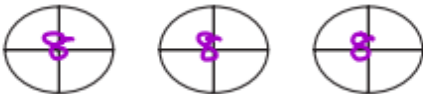

Useful for students to see F P1




Foundation Paper 1 Q1 a



- 1 The incomplete pictogram gives information about the flavour and number of ice creams sold at Pradeep's cafe one Saturday morning.

Flavour	Number of ice creams
Vanilla	 18
Strawberry	 24
Chocolate	 20

Key:  represents 8 ice creams

20 chocolate ice creams were sold on Saturday morning.

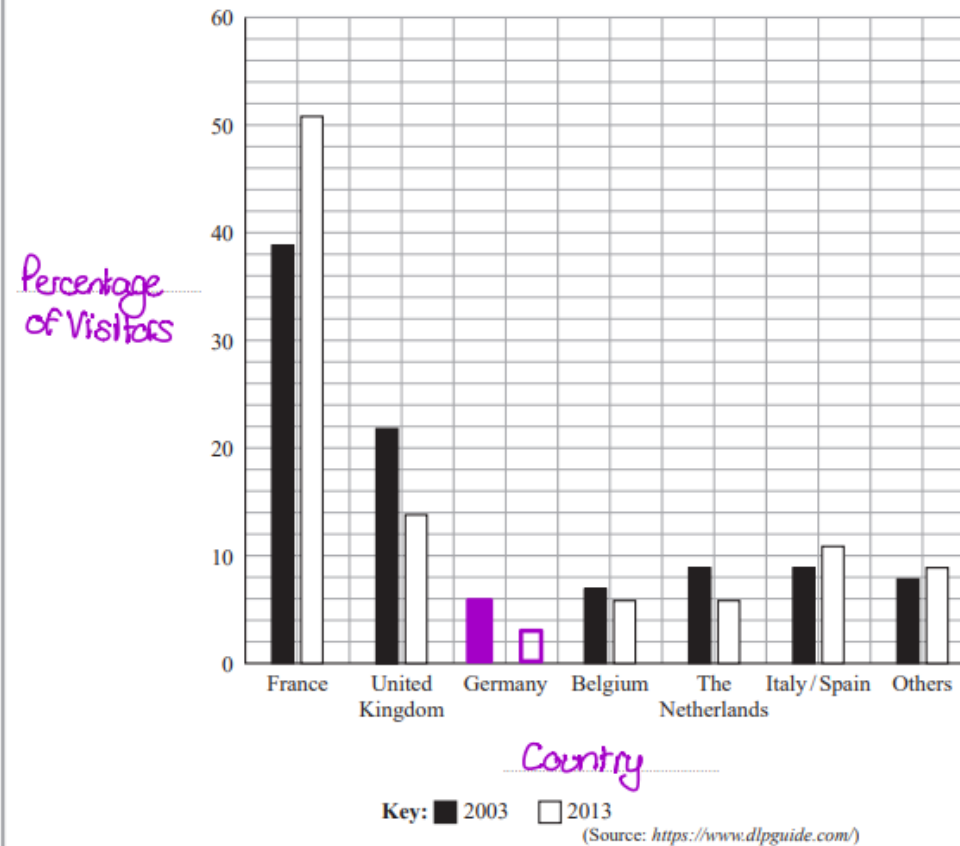
(a) Complete the pictogram for the number of chocolate ice creams sold.

(1)

Foundation Paper 1 Q2 a ii



2 The incomplete multiple bar chart gives information about the home country of visitors to a theme park in 2003 and in 2013



In 2003, 39% of the visitors were from France.

- (a) (i) On the multiple bar chart complete the label for the horizontal axis. (the x axis) (1)
- (ii) On the multiple bar chart complete the label for the vertical axis. (the y axis) (1)

Foundation Paper 1 Q2 d



John is investigating how the total number of visitors to the theme park has changed from 2003 to 2013

- (d) Comment on whether or not it is appropriate to use this multiple bar chart for his investigation.

eg. It is not appropriate as the graph only shows percentages and not any visitor numbers

(2)

(Total for Question 2 is 9 marks)

Foundation Paper 1 Q3 a i



- 3 The tables show information about the number of episodes and viewing figures for two television programs, Emmerdale and Eastenders, for the years 2015 to 2018

<u>Emmerdale</u>	Total number of episodes	Highest viewing figure (millions)	Lowest viewing figure (millions)
Year			
2015	291	6.53	4.04
2016	308	8.03	4.95
2017	302	7.54	5.01
2018	119	7.72	5.72

<u>Eastenders</u>	Total number of episodes	Highest viewing figure (millions)	Lowest viewing figure (millions)
Year			
2015	209	9.87	5.43
2016	210	9.47	4.83
2017	209	8.41	4.19
2018	206	7.81	4.56

(Source: <https://eastenders.fandom.com/wiki/EastEndersWiki>
https://emmerdale.fandom.com/wiki/Emmerdale_Wiki)

- (a) (i) In which of these years did Eastenders have its greatest number of episodes?

2016

(1)

Foundation Paper 1 Q4 a



- 4 A basketball team played 9 matches at the start of a season.

The total number of points they scored in each match is listed below.

✓ 80	✓ 64	✓ 87	✓ 64	✓ 42	✓ 81	✓ 89	✓ 138	✓ 68
42	64	64	68	80	81	87	89	138

Here are some words used to describe data.

grouped

discrete

categorical

continuous

- (a) Select a word from the list to complete the sentence.

The total number of points scored in a match is an example

of discrete data.

(1)

Foundation Paper 1 Q4 c



(c) Give one advantage of using the median to summarise this data.

eg. it is not affected by outliers
(or it is easy to calculate)

(1)

Foundation Paper 1 Q6 b



- 6 The manager of a gym is reviewing the current opening times of the gym.
The manager thinks that if the gym is open for more hours it will affect the number of people using the gym.

(a) Suggest a hypothesis that the manager could use.

eg. The longer the gym is open, the more people will use it.

(1)

The manager wants to get the opinions of the people who have a membership at the gym by giving them a questionnaire.

The manager obtains a numbered list of the 1500 people with a membership and decides to take a sample of 10% of the gym members.

The manager chooses the person who is numbered 0004 as the random starting point on the list and then picks every 20th person.

(b) Name the sampling method that the manager plans to use.

Systematic
(1)

Foundation Paper 1 Q6 f



Following the full survey the manager concludes that if the gym is open for 24 hours a day it will not affect the number of people using the gym.

- (f) Give a reason why it would also be appropriate for the manager to find the opinions of people who do **not** have a gym membership.

eg opening longer hours may encourage them to become members.

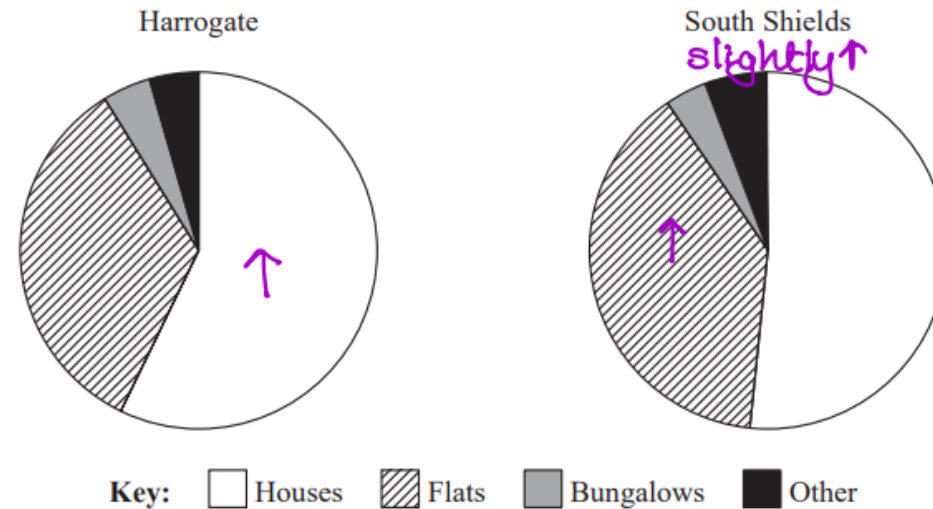
(1)

(Total for Question 6 is 11 marks)

Foundation Paper 1 Q7 a



- 7 The pie charts give information about the different types of properties that were for sale in July 2020 in Harrogate and in July 2020 in South Shields.



(Source: www.rightmove.co.uk)

- (a) Compare the proportion of different types of properties for sale in Harrogate in July 2020 with the proportion of different types of properties for sale in South Shields in July 2020

eg Harrogate had a higher proportion of houses than South Shields.

South Shields had a higher proportion of Flats than Harrogate.

Foundation Paper 1 Q8 c



Emily plans to use a frequency polygon to represent the lengths of the fjords.

- (c) Discuss whether or not a frequency polygon would be an appropriate diagram to use.

eg. Yes it is appropriate for frequency polygons with
or continuous data

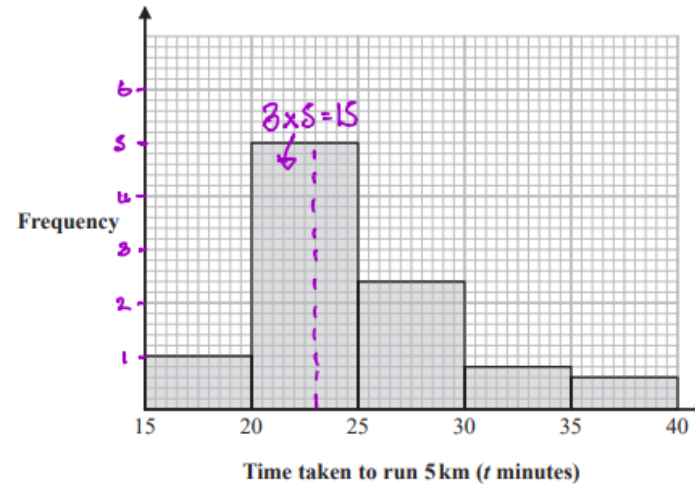
No, because the intervals are not consistent sizes

(2)

Foundation Paper 1 Q10 a (Common Q)



- 10 The time taken, in minutes, for some runners to complete a 5 km run was recorded. The incomplete histogram and incomplete grouped frequency table give information about the times taken, in minutes, for these runners to complete the 5 km run.



Time taken to run 5 km (t minutes)	Frequency
$15 < t \leq 20$	5
$20 < t \leq 25$	25
$25 < t \leq 30$	12
$30 < t \leq 35$	4
$35 < t \leq 40$	3

(Source: www.parkrun.org.uk)

- (a) Use the information in the histogram to complete the table.

Foundation Paper 1 Q10 c (Common Q)



(c) Identify and interpret the skew shown on the histogram.

There is a positive skew.
eg. More than half of the runners take less than the
mean time to complete the 5km

(2)

(Total for Question 10 is 6 marks)

Foundation Paper 1 Q11 a

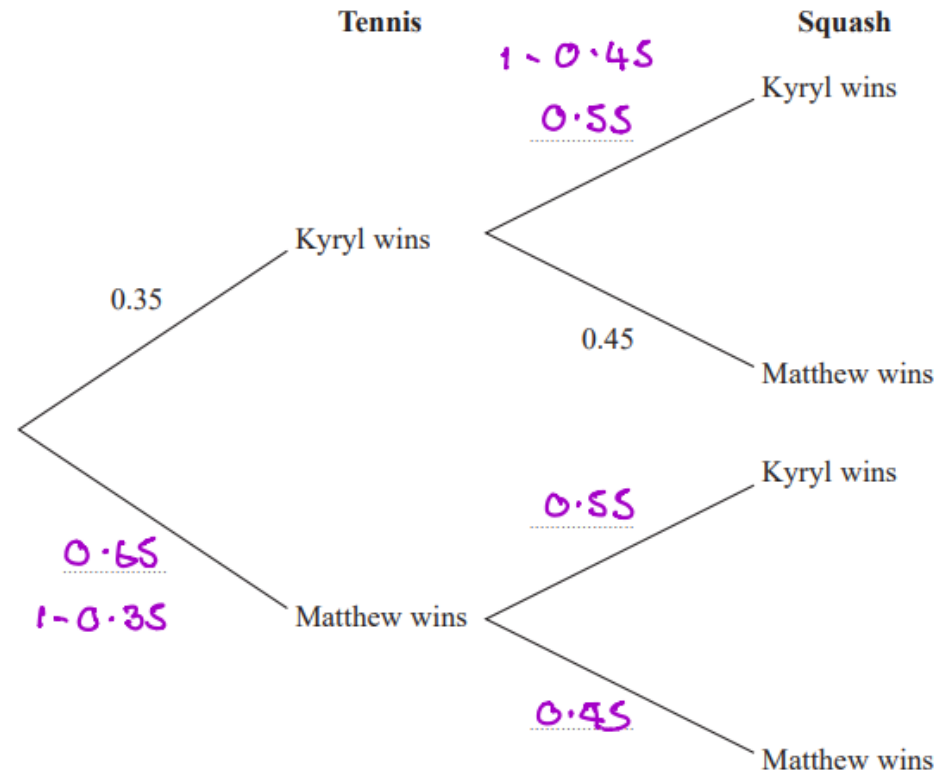


- 11 Kyryl and Matthew play against each other in a game of tennis and a game of squash. In each game either Kyryl or Matthew wins.

The probability that Kyryl wins the game of tennis is 0.35

The probability that Matthew wins the game of squash is 0.45

- (a) Complete the tree diagram to show this information.



(2)

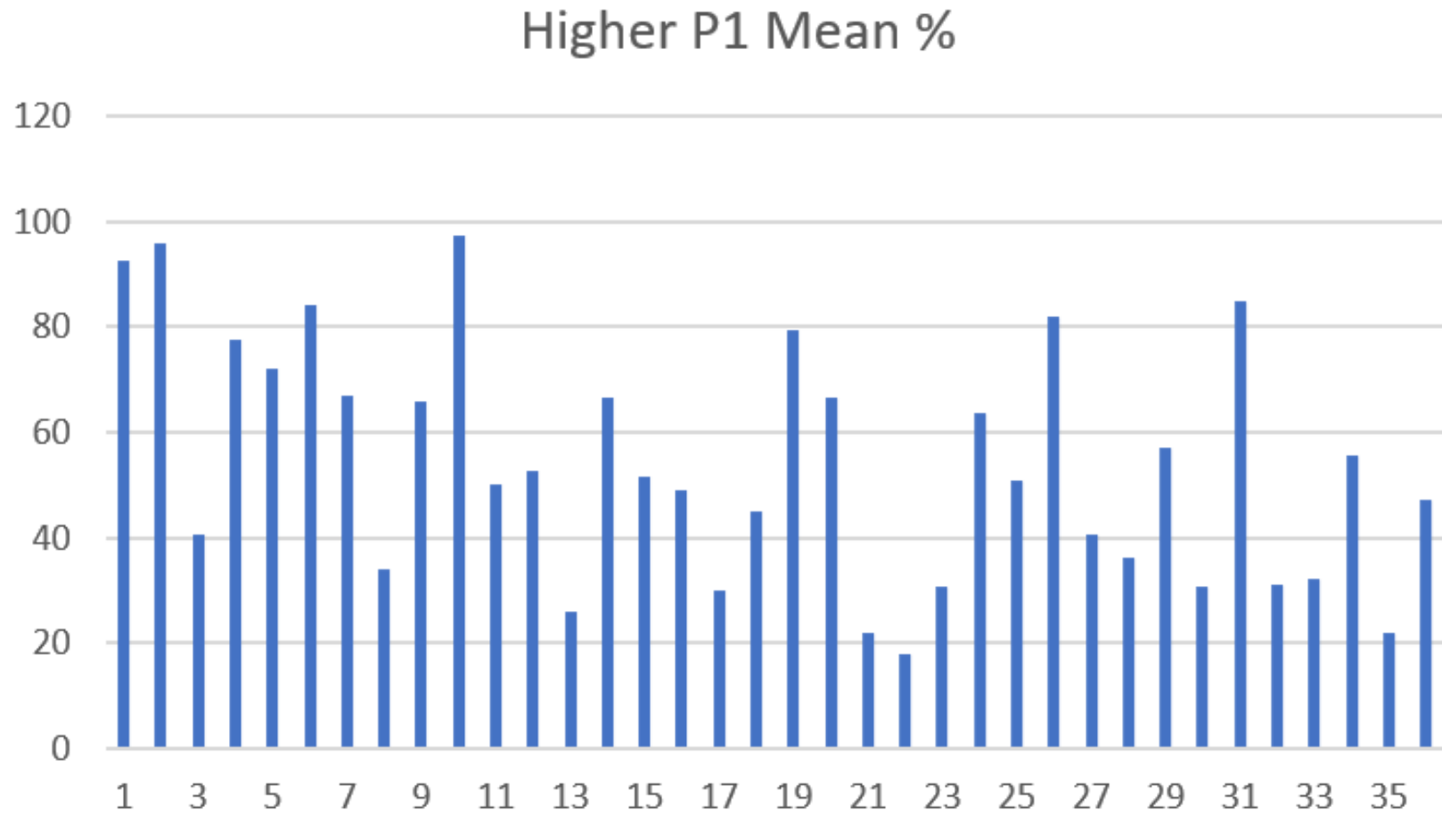
- (b) Matthew says that the probability of him winning both games is greater than the

Higher Paper 1



Q	Part	Position in paper	Performance	Difference	Skill tested	Mean score	Max score	Mean %	9	8	7	6	5	4	3	U	ALL
Q01	ai	1	3	-2	Rates of change over time - using	1.85	2	93	1.98	1.96	1.93	1.89	1.81	1.63	1.28	0.97	1.85
Q01	aii	2	2	0	Rates of change over time - using	0.96	1	96	0.99	0.98	0.98	0.97	0.96	0.90	0.85	0.72	0.96
Q01	b	3	25	-22	Rates of change over time - using	0.81	2	41	1.08	0.87	0.83	0.77	0.77	0.68	0.65	0.42	0.81
Q02	a	4	8	-4	Tabulation - Reading	1.55	2	78	1.95	1.86	1.74	1.59	1.37	1.07	0.75	0.38	1.55
Q02	bi	5	9	-4	Arithmetic mean	2.16	3	72	2.90	2.79	2.61	2.27	1.77	1.07	0.43	0.19	2.16
Q02	bii	6	5	1	Arithmetic mean	0.84	1	84	0.99	0.98	0.94	0.90	0.76	0.58	0.39	0.13	0.84
Q02	biii	7	10	-3	Arithmetic mean	0.67	1	67	0.88	0.81	0.76	0.70	0.59	0.44	0.25	0.02	0.67
Q02	c	8	17	-19	Select and justify appropriate diagrams	0.68	2	34	1.27	1.03	0.82	0.67	0.43	0.22	0.09	0.02	0.68
Q03		9	13	-4	Population pyramid	3.30	5	66	4.69	4.31	3.84	3.33	2.61	1.91	1.24	0.48	3.30
Q04	a	10	1	9	Histograms- using	1.95	2	98	1.99	1.97	1.96	1.95	1.95	1.92	1.89	1.58	1.95
Q04	b	11	20	-9	Histograms - using	1.00	2	50	1.80	1.61	1.31	0.96	0.55	0.25	0.16	0.06	1.00
Q04	c	12	17	-5	Skewness	1.05	2	53	1.79	1.57	1.32	1.05	0.65	0.36	0.20	0.05	1.05
Q05	ai	13	33	-20	Reasons for employing judgement sampling	0.26	1	26	0.65	0.45	0.32	0.21	0.11	0.06	0.03	0.00	0.26
Q05	aii	14	12	2	Reliability and validity	1.33	2	67	1.76	1.68	1.51	1.35	1.16	0.84	0.48	0.20	1.33
Q05	b	15	18	-3	Questionnaires	1.03	2	52	1.65	1.42	1.22	0.97	0.78	0.55	0.31	0.16	1.03
Q05	c	16	21	-5	Problems with collected data	0.98	2	49	1.31	1.16	1.06	0.99	0.86	0.74	0.51	0.23	0.98
Q05	d	17	32	-15	Systematic sampling	0.90	3	30	1.96	1.46	1.11	0.80	0.49	0.26	0.10	0.02	0.90
Q05	e	18	23	-5	Interviews v questionnaires	0.45	1	45	0.74	0.62	0.49	0.42	0.36	0.29	0.19	0.10	0.45
Q06	a	19	7	12	Identify trends in data	1.59	2	80	1.96	1.92	1.84	1.67	1.38	1.03	0.65	0.37	1.59
Q06	b	20	11	9	Interpret seasonal trends in context	2.00	3	67	2.66	2.46	2.28	2.04	1.78	1.26	0.51	0.08	2.00
Q06	ci	21	34	-13	Interpret seasonal trends in context	0.44	2	22	1.63	1.12	0.47	0.16	0.03	0.01	0.00	0.00	0.44
Q06	cii	22	36	-14	Interpret seasonal trends in context	0.18	1	18	0.73	0.46	0.17	0.05	0.01	0.00	0.00	0.00	0.18
Q06	di	23	30	-7	Use trends to make predictions	0.61	2	31	1.90	1.55	0.81	0.25	0.05	0.01	0.00	0.00	0.61
Q06	dii	24	14	10	Use trends to make predictions	1.27	2	64	1.94	1.83	1.58	1.26	0.93	0.55	0.20	0.02	1.27
Q07	a	25	19	6	Sources of data	0.51	1	51	0.76	0.64	0.58	0.51	0.42	0.33	0.22	0.09	0.51
Q07	b	26	6	20	Hypothesis testing	0.82	1	82	0.99	0.97	0.93	0.86	0.75	0.53	0.28	0.12	0.82
Q07	c	27	24	3	Select and justify appropriate diagrams	1.22	3	41	2.48	2.09	1.67	1.09	0.56	0.25	0.07	0.05	1.22
Q07	d	28	26	2	Comparative pie chart	0.72	2	36	1.90	1.60	1.01	0.44	0.12	0.02	0.01	0.00	0.72
Q07	e	29	15	14	Comparative pie chart	0.57	1	57	0.82	0.77	0.69	0.57	0.43	0.31	0.14	0.05	0.57
Q08	a	30	31	-1	Sample means a	0.61	2	31	0.99	0.77	0.63	0.58	0.50	0.41	0.27	0.06	0.61
Q08	b	31	4	27	Use action and warning lines i	1.70	2	85	1.94	1.87	1.82	1.73	1.64	1.43	1.01	0.41	1.70
Q08	c	32	29	3	Use action and warning lines i	1.55	5	31	4.19	3.47	2.21	0.95	0.25	0.07	0.03	0.00	1.55
Q08	d	33	28	5	Use action and warning lines i	0.32	1	32	0.59	0.45	0.35	0.28	0.23	0.19	0.19	0.14	0.32
Q09	a	34	16	18	Tree diagrams	2.23	4	56	3.58	3.07	2.62	2.17	1.65	1.10	0.69	0.49	2.23
Q09	b	35	35	0	Characteristics of a binomial distribution	0.87	4	22	3.26	2.07	1.01	0.35	0.09	0.03	0.01	0.00	0.87
Q10		36	22	14	Use means and standard deviation to standardise and interpret data	2.83	6	47	5.49	5.06	4.31	2.56	0.95	0.33	0.12	0.07	2.83
						41.81	80	52	68.19	59.70	49.73	39.31	29.75	21.63	14.20	7.68	41.81

Useful for students to see H P1



Higher Paper 1 Q1 a ii



- 1 The table gives the total labour force and the unemployment rate for the UK in 2017 and in 2018

Year	Total labour force (million)	Unemployment rate (%)
2017	33.80	4.38
2018	32.41 ↓	4.00 ↓
2019	34.04	3.79

(Source: www.economicsonline.co.uk and www.ons.gov.uk)

In 2019 the total number of unemployed people was 1.29 million.

$$\text{Unemployment rate} = \frac{\text{Number of unemployed people}}{\text{Total labour force}} \times 100$$

- (a) (i) Using the formula above, work out the unemployment rate for the UK in 2019
Give your answer correct to 2 decimal places and write your answer in the table.

(2)

$$\frac{1.29}{34.04} \times 100 = 3.789... \\ \text{so } 3.79$$

- (ii) Using your answer to part (a)(i), what conclusion can be drawn about the unemployment rate in the UK between 2017 and 2019?

The unemployment rate in the UK is decreasing.

(1)

Higher Paper 1 Q1 b



Bob says, without doing any calculations, that the total number of people unemployed decreased from 2017 to 2018

(b) Using the data in the table, assess Bob's claim.

Bob is correct.

eg Both the total workforce and the unemployment rate is lower in 2018 than in 2017.

(2)

Higher Paper 1 Q2 c



Emily plans to use a frequency polygon to represent the lengths of the fjords.

(c) Discuss whether or not a frequency polygon would be an appropriate diagram to use.

eg. Yes it is appropriate for frequency polygons with
or continuous data

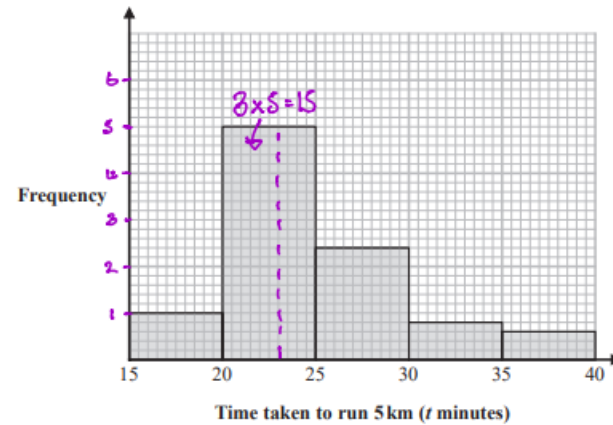
No, because the intervals are not consistent sizes

(2)

Higher Paper 1 Q4 a (Common Q)



- 4 The time taken, in minutes, for some runners to complete a 5 km run was recorded. The incomplete histogram and incomplete grouped frequency table give information about the times taken, in minutes, for these runners to complete the 5 km run.



Time taken to run 5 km (t minutes)	Frequency
$15 < t \leq 20$	5
$20 < t \leq 25$	25
$25 < t \leq 30$	12
$30 < t \leq 35$	4
$35 < t \leq 40$	3

$$2.4 \times 5$$

(Source: www.parkrun.org.uk)

- (a) Use the information in the histogram to complete the table.

(2)

Higher Paper 1 Q5 a i



- 5 The management of a factory is considering changing the working hours of their employees.

Muhammad and Rose want to get the views of the employees in the factory.

Employees in the factory work on the production line or in the warehouse or in the office.

20 employees work on the production line.

15 employees work in the warehouse.

25 employees work in the office.

Muhammad plans to use a questionnaire.

He plans to take a sample of the employees and ask them the questions on his questionnaire.

For his sample, he decides to ask all of the employees who work on the production line.

- (a) (i) Name this sampling technique.

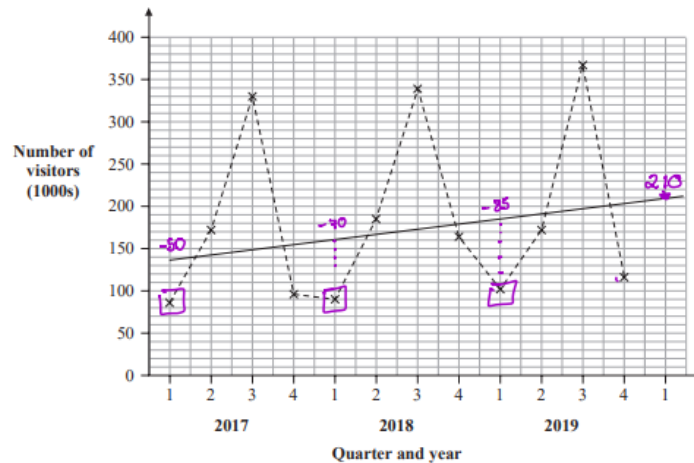
cluster

(1)

Higher Paper 1 Q6 c ii



- 6 The time series graph gives information about the numbers, in thousands, of visitors to Canada from the UK for each quarter for the years 2017 to 2019



(Source: www.ons.gov.uk)

A trend line has been drawn on the time series graph.

- (a) Describe and interpret the trend shown by the graph.

There is an increasing trend which means the number of visitors to Canada from the UK is increasing over time.

(2)

- (b) Discuss any seasonal variation shown by the graph and interpret one in context. Do **not** do any calculations.

eg The lowest quarter every year is quarter 1, which means lower visitors in Q1 from the UK to Canada.

The greatest quarter every year is quarter 3, which means more visitors in Q3 from the UK to Canada.

(3)

- (c) (i) Work out the mean seasonal variation for Quarter 1

Give your answer correct to 1 decimal place. *see graph.*

$$\frac{-50 + (-70) + (-85)}{3}$$

$$= -68.33...$$

- 68.3 thousand
(2)

- (ii) Interpret your answer to part (c)(i) in context.

There are, on average, 68.3 thousand fewer visitors in Q1 compared to the trend.

(1)

- (d) (i) Use your answer to part (c)(i) and the time series graph to predict the number of visitors to Canada from the UK in Quarter 1 of 2020. You must show your working.

$$210 + -68.3$$

$$= 141.7$$

141.7 thousand
(2)

- (ii) Discuss the reliability of using your answer to part (c)(i) to predict the number of visitors to Canada from the UK in Quarter 1 of 2025. Give a reason for your answer.

eg. it may not be reliable because the trend may not continue.

(2)

(Total for Question 6 is 12 marks)

Higher Paper 1 Q7 b



- 7 Roberta is investigating how the ages of brides getting married in the UK has changed from 2003 to 2013
She collects official data from the internet using the website 'Office for National Statistics'.

(a) Explain why this website will give reliable data.

eg The ONS has quality assurance standards.

(1)

Roberta wrote the following hypothesis before she collected her data,

~~Has the age of brides increased between 2003 and 2013?~~

(b) Explain why it is not appropriate to use this as a hypothesis.

it is written as a question. A hypothesis needs to be a statement.

(1)

Higher Paper 1 Q8 b



- 8 Faiza owns a factory that makes footballs. The production line is set up to make footballs with a diameter of 219 mm. As a quality control check, random samples are taken and the mean diameter of each sample is calculated.

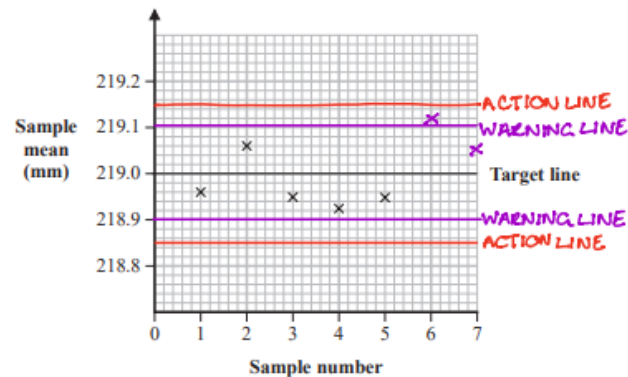
The sample means should be normally distributed with a mean of 219 mm and a standard deviation of 0.05 mm.

- (a) Would the sample means or the population values be more consistent? Give a reason for your answer.

The sample means would be more consistent because the standard deviation is smaller.

(2)

Faiza draws a control chart showing the first 5 samples.



Sample 6 and sample 7 have the following sample means.

	Sample 6	Sample 7
Sample mean (mm)	219.12	219.05

- (b) Plot these two sample means on the quality control chart.

(2)

- (c) By completing the control chart, determine what actions, if any, Faiza should have taken based on the information given after each of these last two samples. You must justify your answer.

Sample 6 is between the warning and action line so another sample should be taken

Sample 7 is within the warning lines so the machine can continue.

A small, bright yellow five-pointed star with a friendly face, featuring two large, dark eyes and a small, curved smile. It is positioned in the top right corner of the page.

- The probability that Peter drops his children off on time on a Monday morning is three times the probability that he drops them off late on a Monday morning.

The probability that Peter drops them off late on a Tuesday morning given that he has dropped them off late on a Monday morning is 0.3

Monday **Tuesday**

On time 0.75 0.8 On time

Late 0.25 0.2 Late

On time 0.7 $1-0.8$ On time

Late 0.3 $1-0.3$ Late

$$P(0,0) = 0.6 = 0.75 \times 2$$

$$x = \frac{0.6}{0.75} = 0.8$$

Higher Paper 1 Q9 b



- 9 Peter drops his children off at a breakfast club every Monday morning and every Tuesday morning.

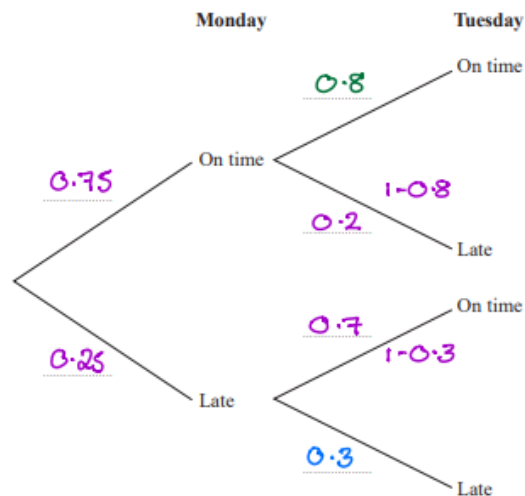
He either drops them off on time or late.

The probability that Peter drops his children off on time on a Monday morning is three times the probability that he drops them off late on a Monday morning. 0.75 0.25

The probability that Peter drops them off on time on a Monday morning **and** on a Tuesday morning is 0.6

The probability that Peter drops them off late on a Tuesday morning given that he has dropped them off late on a Monday morning is 0.3

- (a) Complete the tree diagram below.



$$P(0,0) = 0.6 = 0.75 \times x$$

$$x = \frac{0.6}{0.75} = 0.8$$

(4)

Peter drops his children off at the breakfast club every Monday morning and Tuesday morning over a 4-week period.

Assume that the probability of dropping them off on time on both a Monday morning and a Tuesday morning remains constant and that each week is independent.

- (b) Work out the probability that, over a 4-week period, he drops his children off on time on a Monday morning and on time on a Tuesday morning on 3 or more occasions. $\text{ie } 3 \text{ or } 4$

$$\text{using } P(\text{on time on both days}) = 0.6$$

$$\text{so } P(\text{not on time on both days}) = 0.4$$

$$P(X=3) = 4 \times 0.6^3 \times 0.4$$

$$= 0.3456 \dots$$

$$P(X=4) = 0.6^4$$

$$= 0.1296$$

$$\text{so } 0.3456 \dots + 0.1296$$

$$= 0.4752$$

$$0.475 \text{ (3dp)}$$

(4)

(Total for Question 9 is 8 marks)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

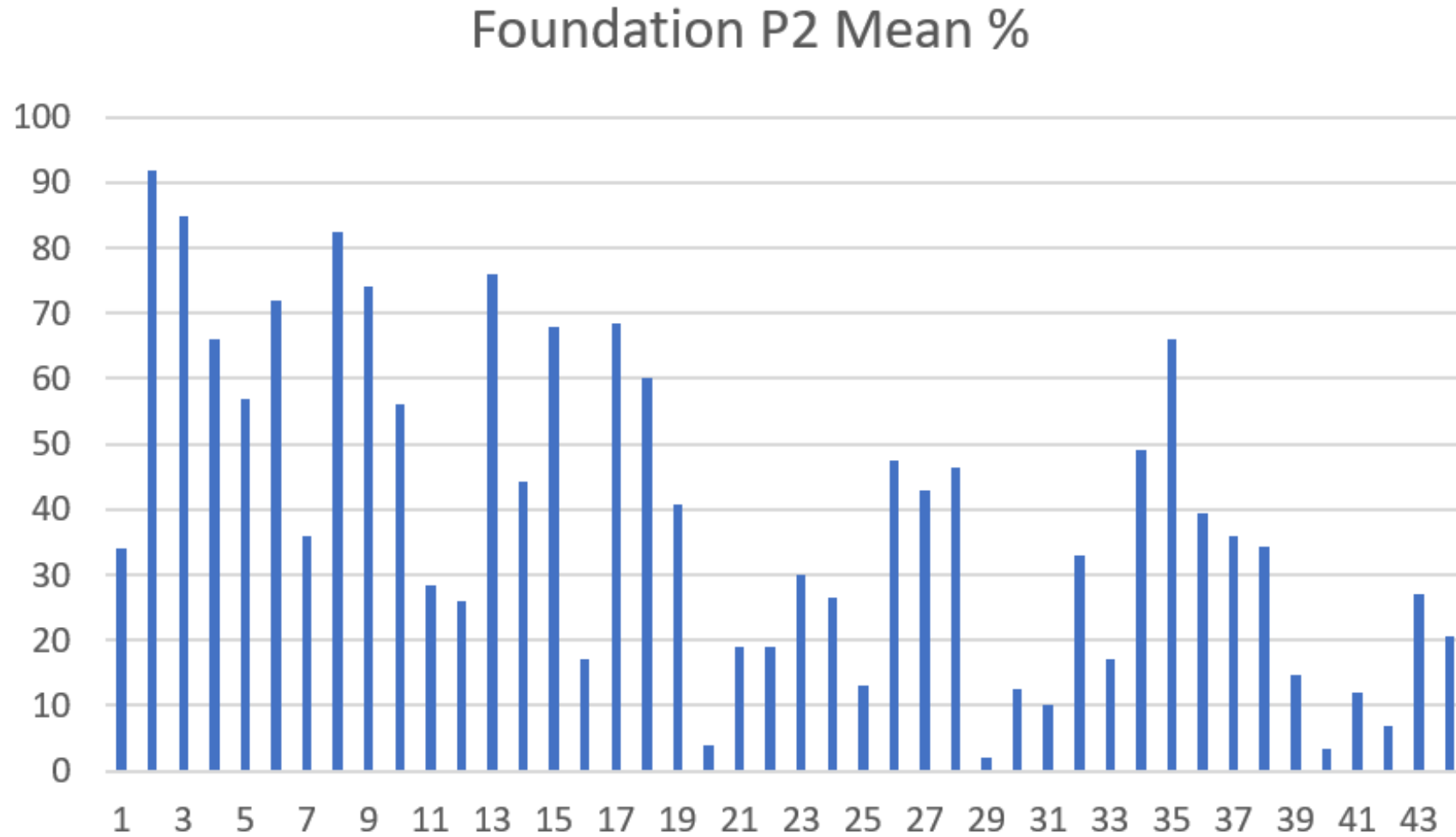
DO NOT WRITE IN THIS AREA

Foundation Paper 2



Q	Part	Position in paper	Performance	Difference	Skill tested	Mean score	Max score	Mean %	5	4	3	2	1	U	ALL
Q01	a	1	24	-23	Probability statements of likelihood	0.34	1	34	0.51	0.41	0.33	0.24	0.15	0.11	0.34
Q01	b	2	1	1	Probability statements of likelihood	0.92	1	92	0.99	0.98	0.97	0.91	0.76	0.41	0.92
Q01	c	3	2	1	Probability scale	0.85	1	85	0.98	0.95	0.89	0.78	0.65	0.31	0.85
Q01	d	4	9	-5	Probability scale	0.66	1	66	0.93	0.83	0.66	0.45	0.30	0.10	0.66
Q02	a	5	12	-7	Bar chart with probabilities	0.57	1	57	0.90	0.79	0.56	0.31	0.11	0.02	0.57
Q02	b	6	6	0	Bar charts - mode from	0.72	1	72	0.95	0.87	0.74	0.57	0.38	0.18	0.72
Q02	c	7	21	-14	Bar charts - interpreting	0.36	1	36	0.60	0.48	0.34	0.21	0.09	0.02	0.36
Q03	a	8	3	5	Tally	1.65	2	83	1.89	1.84	1.71	1.55	1.20	0.51	1.65
Q03	b	9	5	4	Select an appropriate diagram	0.74	1	74	0.88	0.82	0.79	0.69	0.51	0.14	0.74
Q03	c	10	13	-3	Calculate mode	0.56	1	56	0.87	0.73	0.54	0.36	0.19	0.04	0.56
Q03	d	11	27	-16	Calculate median	0.57	2	28	1.24	0.72	0.39	0.22	0.14	0.05	0.57
Q03	e	12	30	-18	Rationale for selecting types of average	0.52	2	26	0.79	0.64	0.52	0.35	0.17	0.05	0.52
Q04	a	13	4	9	Probability from a list	0.76	1	76	0.95	0.92	0.83	0.65	0.31	0.08	0.76
Q04	b	14	17	-3	Compare data sets	1.77	4	44	3.10	2.44	1.67	0.73	0.24	0.03	1.77
Q04	c	15	8	7	Calculate using range	0.68	1	68	0.98	0.90	0.71	0.42	0.20	0.08	0.68
Q05		16	34	-18	Planning data collection	1.02	6	17	2.39	1.36	0.67	0.26	0.10	0.01	1.02
Q06	a	17	7	10	Bar charts - completing	1.37	2	69	1.80	1.63	1.47	1.11	0.58	0.12	1.37
Q06	b	18	11	7	Bar charts - interpreting	0.60	1	60	0.85	0.74	0.60	0.45	0.23	0.08	0.60
Q06	c	19	19	0	Bar charts - interpreting	1.22	3	41	2.13	1.61	1.09	0.66	0.28	0.09	1.22
Q07	a	20	42	-22	Definition of population	0.04	1	4	0.07	0.05	0.03	0.02	0.01	0.00	0.04
Q07	b	21	32	-11	Definition of convenience sample	0.19	1	19	0.40	0.24	0.14	0.07	0.03	0.00	0.19
Q07	c	22	33	-11	Disadvantages of sampling	0.19	1	19	0.45	0.25	0.13	0.06	0.02	0.00	0.19
Q07	d	23	26	-3	Tabulation	0.60	2	30	1.01	0.76	0.57	0.36	0.15	0.03	0.60
Q07	e	24	29	-5	Select an appropriate diagram	0.53	2	27	0.93	0.66	0.47	0.30	0.14	0.03	0.53
Q08	a	25	37	-12	Select an appropriate diagram	0.13	1	13	0.32	0.16	0.07	0.04	0.02	0.00	0.13
Q08	b	26	15	11	Scatter diagrams - completing	0.95	2	48	1.65	1.30	0.90	0.44	0.13	0.01	0.95
Q08	c	27	18	9	Correlation	0.86	2	43	1.48	1.17	0.81	0.42	0.15	0.01	0.86
Q08	d	28	16	12	Determine line of best fit	0.93	2	47	1.65	1.29	0.85	0.40	0.15	0.04	0.93
Q08	e	29	44	-15	Interpret Spearman's rank correlation coefficient	0.02	1	2	0.06	0.02	0.01	0.00	0.00	0.00	0.02
Q08	f	30	38	-8	Interpret Spearman's rank correlation coefficient	0.25	2	13	0.51	0.35	0.21	0.08	0.03	0.00	0.25
Q08	g	31	40	-9	Reliability and validity	0.20	2	10	0.65	0.25	0.06	0.02	0.01	0.00	0.20
Q09	a	32	25	7	Cleaning data	0.66	2	33	1.09	0.90	0.62	0.32	0.11	0.02	0.66
Q09	b	33	35	-2	Select an appropriate diagram	0.17	1	17	0.40	0.22	0.11	0.04	0.01	0.00	0.17
Q10	a	34	14	20	Difference between primary and secondary data	0.49	1	49	0.80	0.67	0.47	0.25	0.10	0.02	0.49
Q10	b	35	10	25	Choropleth map - reading	0.66	1	66	0.92	0.84	0.69	0.47	0.25	0.07	0.66
Q10	c	36	20	16	Choropleth map - interpreting	0.79	2	40	1.32	1.12	0.76	0.35	0.12	0.01	0.79
Q11	a	37	22	15	Definitions for types of data	0.36	1	36	0.58	0.41	0.31	0.26	0.21	0.10	0.36
Q11	b	38	23	15	Box plots - reading	1.03	3	34	2.53	1.51	0.56	0.16	0.03	0.00	1.03
Q11	c	39	36	3	Box plots - comparing	0.74	5	15	2.17	0.98	0.26	0.05	0.01	0.00	0.74
Q11	d	40	43	-3	Use stratification	0.10	3	3	0.37	0.09	0.02	0.01	0.00	0.00	0.10
Q12	a	41	39	2	Determine relative risks	0.36	3	12	0.90	0.50	0.21	0.06	0.02	0.00	0.36
Q12	b	42	41	1	Interpret relative risks	0.07	1	7	0.22	0.09	0.03	0.01	0.00	0.00	0.07
Q13	a	43	28	15	Median from cumulative frequency graph	0.27	1	27	0.55	0.31	0.20	0.14	0.09	0.02	0.27
Q13	b	44	31	13	Cumulative frequency graph - comparing	0.82	4	21	2.28	1.03	0.40	0.13	0.03	0.00	0.82
						27.29	80	34	47.04	34.83	24.37	15.38	8.41	2.79	27.29

Useful for students to see F P2



Foundation Paper 2 Q1 a, b and c

- 1 David has 10 cards each with a single letter on it as shown.



A card is picked at random.

- (a) Underline the word from the list below that best describes the likelihood that the card has a letter A on it.

impossible certain likely evens unlikely

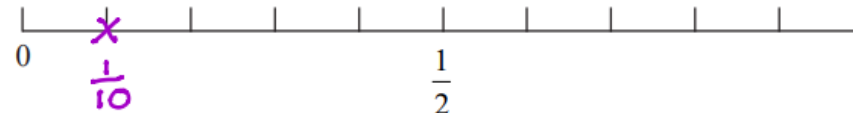
(1)

- (b) Complete this sentence using two different letters.

Cards with the letters B and D are equally likely to be picked.

(1)

- (c) On the probability scale below, mark with a cross (×) the probability that the card has a letter C on it.



(1)

Foundation Paper 2 Q3 d and e



- 3 Ben is researching information about the number of British swimming medals won at the Olympics.

Here are his results, giving the number of British swimming medals won at the Olympics from 1900 to 2016

3	0	7	6	2	4	4	2	8
1	1	2	1	1	1	1	3	5
5	2	1	2	0	2	1	3	6

(Source: www.teamgb.com)

- (a) Fill in the tally chart for Ben's results and complete the frequency column.

Number of Olympic medals won	Tally	Frequency
0		3
1	1	6
2		5
3	1	6
4		2
5		2
6		2
7		1

(2)

- (b) Suggest a suitable diagram that could be used for Ben's results.

e.g. pictogram or bar chart
or pie chart

(1)

- (c) Write down the mode or modes.

1 and 3

(1)

- (d) Work out the median.

27 in total so $\frac{27+1}{2} = 14\text{th value}$

$3+6=9$ $9+5=14$ ✓

2

(2)

Ben wants to use an average to summarise the data.

- (e) Which of the mode or the median would be more appropriate?
Give a reason for your answer.

eg. The median as there is more than one mode

(2)

(Total for Question 3 is 8 marks)



Foundation Paper 2 Q5



- 5 Claire is planning an investigation into the length of time that a learner has to wait for a driving test. She wants to find out about how waiting time varies in different regions of the UK. Here is her plan for data collection, for calculations and for diagrams.

Data collection

Visit a random sample of driving test centres in each region to ask for their waiting time in June.

Calculations

Calculate the average waiting time for each region for June.

Calculate the range of the waiting times for each region for June.

Diagrams

Draw a bar chart showing the average waiting time for each region in June.

Draw a pie chart showing the range of waiting times for each region in June.

Discuss whether Claire's plans for data collection, for calculations and for diagrams are appropriate.

Data collection

eg • taking a random sample means each test centre has an equally likely chance of being selected but it would be quicker, easier and cheaper to collect data from the internet. Data for more than one month should be collected which would make it more appropriate.

Calculations.

Calculating an average is a good way to compare waiting times for each region but she hasn't said which average she'll use. The range is a good way to measure spread so is appropriate.

Diagrams. A bar chart is not appropriate for wait times as it's continuous data and a pie chart is not appropriate either.

Hint: Look at the number of marks (Total for Question 5 is 6 marks) for guidance about the number of points to make.

Foundation Paper 2 Q7 a



7 Chris is a manager at a theme park.

He wants to find out what food options visitors would like to be able to buy in the theme park.

(a) State the population for this investigation.

all visitors to the theme park.

(1)

Foundation Paper 2 Q8e



7 Chris is a manager at a theme park.

He wants to find out what food options visitors would like to be able to buy in the theme park.

(a) State the population for this investigation.

all visitors to the theme park.

(1)

Chris decides that he will take a convenience sample of visitors in the section of the park selling food.

(b) Explain what is meant by a convenience sample.

eg a sample from people who are available at the time.

(1)

(c) Give one disadvantage of using a convenience sample.

It may not be representative or may be biased

(1)

Chris plans to use the data collection sheet below.

Type of food	Tally
Pizza	
Chinese	
Curry	
Fish and chips	

(d) Discuss whether this data collection sheet is appropriate.

You should consider how Chris might use the data and describe any problems he might have when he uses the data collection sheet.

eg. There are a very limited number of choices and there is no 'other' option or an option for people who don't buy food at the theme park so it is not appropriate.

(2)

Chris suggests using a stem and leaf diagram to represent the data that he collects.

(e) Discuss whether or not this would be a suitable diagram to represent his data.

eg. A stem and leaf is not suitable as the data being collected is qualitative and stem and leaf is better suited for numerical data.

(2)

(Total for Question 7 is 7 marks)

Foundation Paper 2 Q10 a, b and c (Common Q)



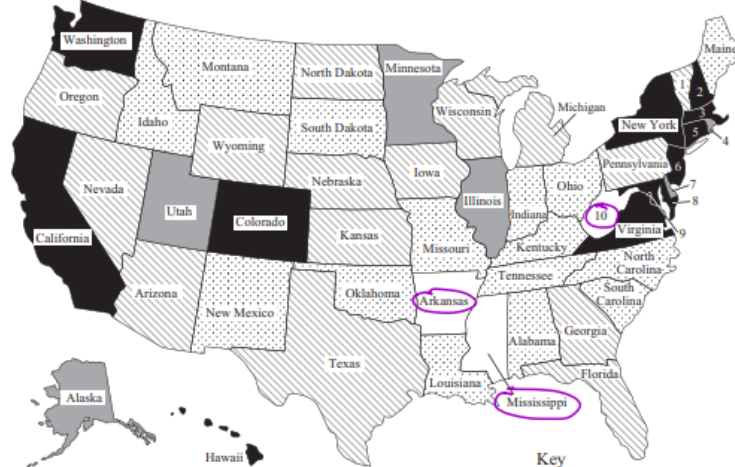
10 Matthew is investigating average household income for different states in the USA.

(a) Give a reason why it is appropriate to use secondary data for this.

eg It would take too long to collect all the data himself.
or It would be very expensive to collect the data himself.
(1)

Matthew creates a choropleth map giving information about the mean household income by state for the USA in 2023

Mean annual household income in \$ thousands.



1. Vermont
2. New Hampshire
3. Massachusetts
4. Rhode Island
5. Connecticut
6. New Jersey
7. Delaware
8. Maryland
9. District of Columbia
10. West Virginia

(Source: worldpopulationreview.com)



(b) Which **three** states have the lowest mean household income?

West Virginia, Arkansas, Mississippi

(1)

Matthew concludes that the mean household incomes are highest on the West coast and the East coast.

(c) Does the choropleth map support this conclusion?
Give a reason for your answer.

W N E
S

eg. In general the map supports this conclusion as the highest mean incomes ie the darkest areas are either on the west or east coast, with the exception of Colorado.
(2)

(Total for Question 10 is 4 marks)

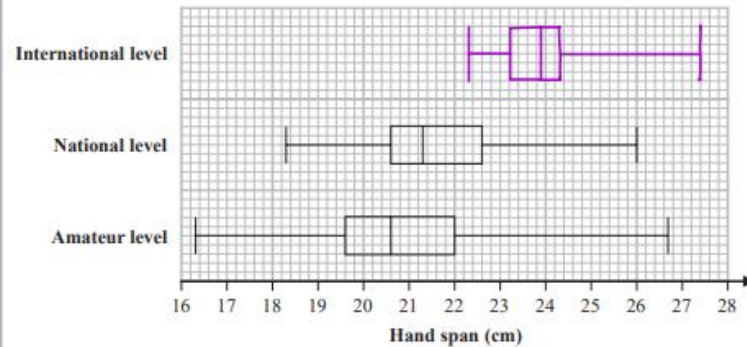


Foundation Paper 2 Q11 d (Common Q)



- 11 Some researchers investigated the hand span, in centimetres, of adult pianists by their level – international, national and amateur.

The box plots below give information about the hand spans for national level and amateur level pianists.



(Source: www.appca.com.au)

- (a) Circle the word in the list below that describes hand span, in centimetres, as a type of data.

qualitative ordinal continuous bivariate

(1)

The table gives information about the hand spans of the international level pianists.

Greatest hand span	27.4 cm
Median hand span	23.9 cm
Lower quartile	23.2 cm
Range	5.1 cm
Interquartile range	1.1 cm

$U.Q. \quad 23.2 + 1.1 = 24.3$

- (b) Using the information in the table, draw on the grid above a box plot for the hand spans of the international level pianists.

(3) ✓

- (c) Compare the three distributions of hand spans.

Give **three** comparisons and interpret **two** of your comparisons.

eg Pianists at International have the highest median which means, on average they have the largest hand spans.

The pianists at International level have the smallest IQR so they have more consistent hand spans.

All three have positive skew, so have more varied hand spans above the median than below the median

(5)

Pavel owns a music shop.

He wants to investigate the keyboard sizes used by pianists with different hand spans. He collects data about the hand spans of the pianists who use his shop.

The table gives information about the number of these pianists with hand spans in each of four size categories.

Hand span (cm)	A (less than 19)	B ($19 \leq \text{span} < 22$)	C ($22 \leq \text{span} < 24$)	D (24 or more)
Number of pianists	24	65	57	14

160

$\frac{24}{160} \times 20 = 3$ $\frac{65}{160} \times 20 = 8.125$ $\frac{57}{160} \times 20 = 7.125$ $\frac{14}{160} \times 20 = 1.75$

Pavel plans to sample 20 of these pianists stratified by hand span size.

- (d) Explain how Pavel can obtain his stratified sample.

You should include details of any calculations he should use.

• By taking 3 from A, 8 from B, 7 from C and 2 from D.

• There should be a random sample, for example all the pianists within each strata could be labelled and using a random number generator to select the appropriate number of pianists.

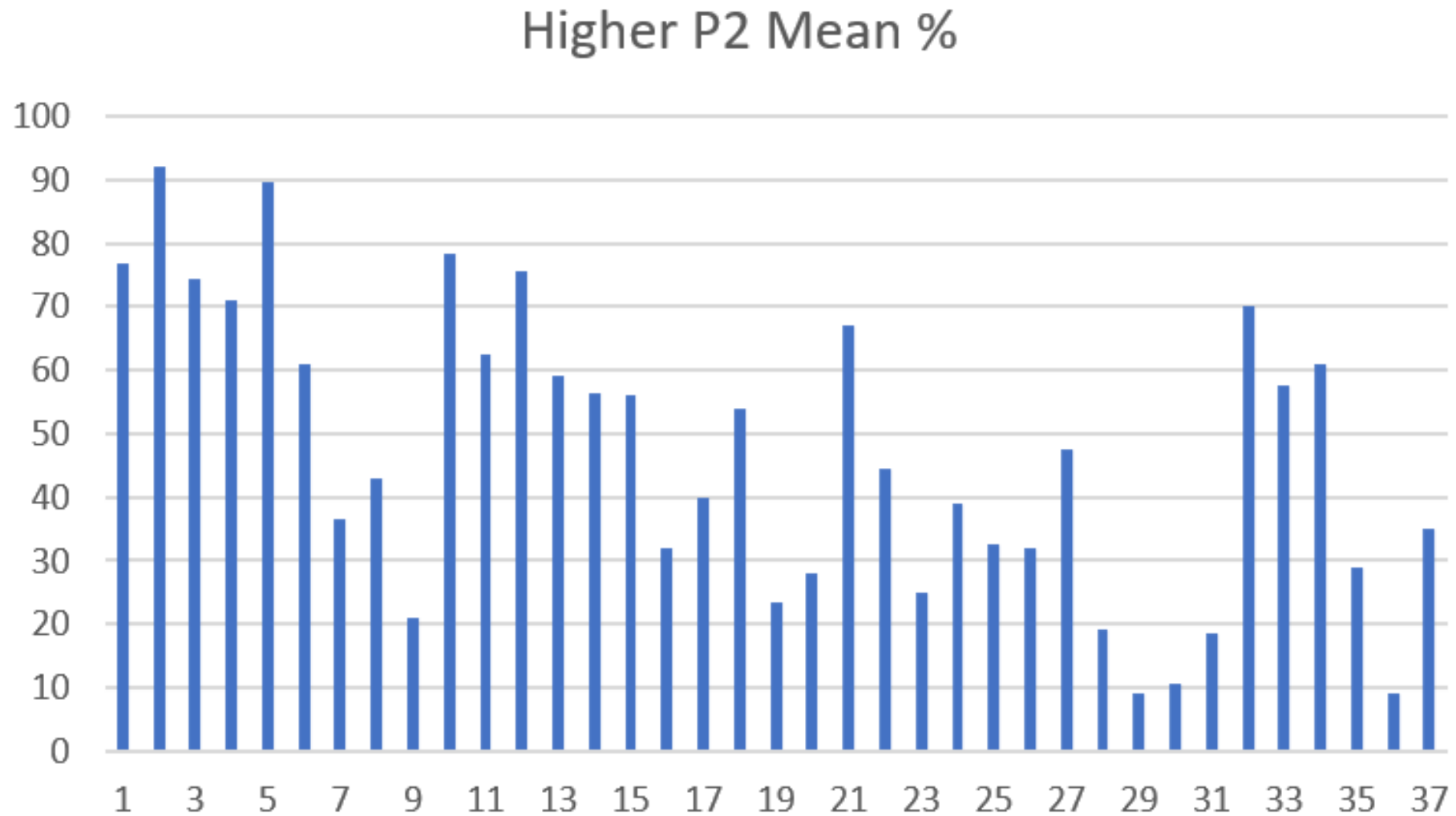
(3)

Higher Paper 2



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
	Q	Part	Position in paper	Performance	Difference	Skill tested	Mean score	Max score	Mean %	9	8	7	6	5	4	3	U	ALL
1																		
2	Q01	a	1	4	-3	Difference between primary and secondary data	0.77	1	77	0.92	0.87	0.86	0.80	0.72	0.56	0.36	0.25	0.77
3	Q01	b	2	1	1	Choropleth map - reading	0.92	1	92	0.99	0.97	0.95	0.94	0.91	0.84	0.70	0.50	0.92
4	Q01	c	3	6	-3	Choropleth map - interpreting	1.49	2	75	1.76	1.66	1.59	1.51	1.40	1.25	0.94	0.44	1.49
5	Q02	a	4	7	-3	Definitions for types of data	0.71	1	71	0.97	0.92	0.83	0.72	0.57	0.43	0.31	0.32	0.71
6	Q02	b	5	2	3	Box plots - reading	2.69	3	90	2.92	2.93	2.88	2.83	2.67	2.11	1.16	0.50	2.69
7	Q02	c	6	11	-5	Box plots - comparing	3.05	5	61	4.48	4.08	3.63	3.12	2.40	1.48	0.54	0.21	3.05
8	Q02	d	7	23	-16	Use stratification	1.10	3	37	2.21	1.90	1.52	1.04	0.45	0.14	0.01	0.00	1.10
9	Q03	a	8	20	-12	Determine relative risks	1.29	3	43	2.77	2.15	1.49	1.04	0.76	0.52	0.24	0.08	1.29
10	Q03	b	9	32	-23	Interpret relative risks	0.21	1	21	0.40	0.29	0.25	0.20	0.17	0.09	0.02	0.02	0.21
11	Q04	a	10	3	7	Cumulative frequency- drawing	1.57	2	79	1.91	1.83	1.72	1.60	1.42	1.18	1.03	0.62	1.57
12	Q04	b	11	10	1	Cumulative frequency - using	1.25	2	63	1.89	1.77	1.56	1.28	0.92	0.46	0.20	0.10	1.25
13	Q05	a	12	5	7	Stem and leaf diagram	2.27	3	76	2.71	2.51	2.41	2.27	2.12	1.89	1.54	0.94	2.27
14	Q05	b	13	13	0	Calculate median	0.59	1	59	0.87	0.76	0.69	0.60	0.48	0.33	0.19	0.09	0.59
15	Q05	c	14	15	-1	Calculate interquartile range	1.13	2	56	1.76	1.58	1.42	1.16	0.82	0.39	0.10	0.02	1.13
16	Q05	d	15	16	-1	Compare data sets using appropriate measures	1.68	3	56	2.63	2.42	2.12	1.77	1.17	0.50	0.14	0.04	1.68
17	Q05	e	16	26	-10	Comment on appropriateness	0.32	1	32	0.64	0.52	0.41	0.28	0.16	0.09	0.04	0.01	0.32
18	Q05	f	17	21	-4	Limitations of data sources	0.40	1	40	0.72	0.60	0.48	0.38	0.28	0.13	0.02	0.01	0.40
19	Q06	a	18	17	1	Determine Spearman's rank correlation coefficient	3.23	6	54	5.23	4.77	4.24	3.30	2.11	0.88	0.30	0.27	3.23
20	Q06	b	19	31	-12	Extraneous variables	0.47	2	24	1.04	0.76	0.56	0.43	0.27	0.11	0.03	0.00	0.47
21	Q06	c	20	29	-9	Controlling extraneous variables	0.28	1	28	0.79	0.55	0.35	0.19	0.08	0.02	0.00	0.00	0.28
22	Q07	a	21	9	12	Use index numbers in context	1.34	2	67	1.91	1.87	1.71	1.39	0.94	0.54	0.25	0.08	1.34
23	Q07	b	22	19	3	Use index numbers in context	0.89	2	45	1.41	1.18	1.01	0.86	0.75	0.45	0.14	0.01	0.89
24	Q07	c	23	30	-7	Calculate price index	0.50	2	25	1.80	1.27	0.58	0.19	0.04	0.01	0.01	0.00	0.50
25	Q07	d	24	22	2	Use appropriate measure of central tendency	0.39	1	39	0.73	0.61	0.48	0.34	0.26	0.18	0.09	0.03	0.39
26	Q07	e	25	25	0	Calculate												

Useful for students to see H P2



Higher Paper 2 Q1 b (Common Q)



Answer ALL questions.
Write your answers in the spaces provided.
You must write down all the stages in your working.

1 Matthew is investigating average household income for different states in the USA.

(a) Give a reason why it is appropriate to use secondary data for this.

eg It would take too long to collect all the data himself.
or It would be very expensive to collect the data himself.

(1)

Matthew creates a choropleth map giving information about the mean household income by state for the USA in 2023

Mean annual household income in \$ thousands.

Key

- 1. Vermont
- 2. New Hampshire
- 3. Massachusetts
- 4. Rhode Island
- 5. Connecticut
- 6. New Jersey
- 7. Delaware
- 8. Maryland
- 9. District of Columbia
- 10. West Virginia

(Source: worldpopulationreview.com)

(b) Which three states have the lowest mean household income?

West Virginia, Arkansas, Mississippi

(1)

Matthew concludes that the mean household incomes are highest on the West coast and the East coast.

(c) Does the choropleth map support this conclusion?

Give a reason for your answer.

eg In general the map supports this conclusion as the highest mean incomes ie the darkest areas are either on the west or east coast, with the exception of Colorado.

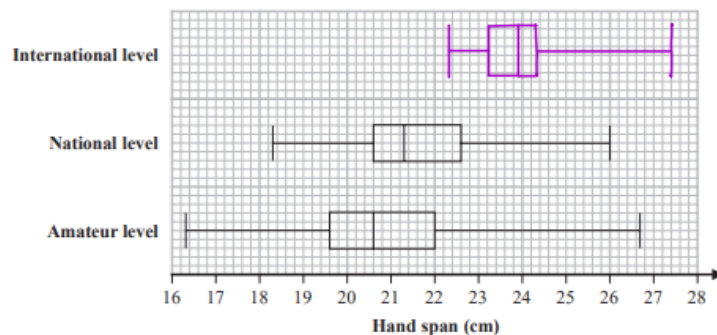
(2)

(Total for Question 1 is 4 marks)

Higher Paper 2 Q2 b and d (Common Q)

- 2 Some researchers investigated the hand span, in centimetres, of adult pianists by their level – international, national and amateur.

The box plots below give information about the hand spans for national level and amateur level pianists.



(Source: www.appca.com.au)

- (a) Circle the word in the list below that describes hand span, in centimetres, as a type of data.

qualitative ordinal continuous bivariate

(1)

The table gives information about the hand spans of the international level pianists.

Greatest hand span	27.4 cm
Median hand span	23.9 cm
Lower quartile	23.2 cm
Range	5.1 cm
Interquartile range	1.1 cm

$U.Q. \quad 23.2 + 1.1 = 24.3$

- (b) Using the information in the table, draw on the grid above a box plot for the hand spans of the international level pianists.

(3)

- (c) Compare the three distributions of hand spans.

Give **three** comparisons and interpret **two** of your comparisons.

eg Pianists at International have the highest median which means, on average they have the largest hand spans.

The pianists at International level have the smallest IQR so they have more consistent hand spans.

All three have positive skew, so have more varied hand spans above the median than below the median

(5)

Pavel owns a music shop.

He wants to investigate the keyboard sizes used by pianists with different hand spans. He collects data about the hand spans of the pianists who use his shop.

The table gives information about the number of these pianists with hand spans in each of four size categories.

Hand span (cm)	A (less than 19)	B (19 ≤ span < 22)	C (22 ≤ span < 24)	D (24 or more)
Number of pianists	24	65	57	14

$\frac{24}{160} \times 20 = 3$ $\frac{65}{160} \times 20 = 8.125$ $\frac{57}{160} \times 20 = 7.125$ $\frac{14}{160} \times 20 = 1.75$

Pavel plans to sample 20 of these pianists stratified by hand span size.

- (d) Explain how Pavel can obtain his stratified sample.

You should include details of any calculations he should use.

• By taking 3 from A, 8 from B, 7 from C and 2 from D.

• There should be a random sample, for example all the pianists within each strata could be labelled and using a random number generator to select the appropriate number of pianists

Higher Paper 2 Q3 b (Common Q)



- 3 Khatia organises two different training courses, Course A and Course B, to help people to learn to type. She wants to compare the two different courses to see which is better. At the end of each course the people are given a skills test.

The table shows the number of participants who passed and failed the skills test for each of the two courses.

	Passed	Failed	Total
Course A	35	15	50
Course B	48	32	80

Risk of failing
15/50
32/80

- (a) Find the relative risk of failing the skills test having taken Course A compared to Course B.

$$\frac{15}{50} \div \frac{32}{80} = 0.75$$

0.75

(3)

- (b) Give an interpretation of your answer to part (a).

eg. The risk of failing the skills test on Course A is lower than the risk of failing on Course B.

(1)

(Total for Question 3 is 4 marks)

Higher Paper 2 Q9 c



- 9 Researchers wanted to find a method to predict the height of ancient Egyptians based upon bones from their skeletons.

They calculated the Pearson's product moment correlation coefficient between the length of particular bones and the height of the skeleton for some male skeletons and for some female skeletons.

They also found regression equations for the relationship between bone length (x centimetres) and height (y centimetres) for each of these bones in male skeletons and in female skeletons.

The table gives information about these product moment correlation coefficients and these regression equations.

	Product moment correlation coefficient	Regression equation
Males		
Femur	0.826	$y = 2.257x + 63.93$
Tibia	0.850	$y = 2.554x + 69.21$
Humerus	0.656	$y = 2.594x + 83.85$
Radius	0.649	$y = 2.641x + 100.91$
Females		
Femur	0.891	$y = 2.340x + 56.99$
Tibia	0.938	$y = 2.699x + 61.08$
Humerus	0.806	$y = 2.827x + 70.94$
Radius	0.580	$y = 2.509x + 96.73$

(Source: www.semanticscholar.org)

- (a) Which bone measurement would you recommend using to estimate the height of an ancient Egyptian?
Give a reason for your answer.

The tibia length because the PMCC is the highest for both males and females.

- (b) Interpret in context the figure 2.594 in the regression equation for the humerus of male skeletons.

It means that for every additional cm of length of the humerus the height is estimated to increase by 2.594 cm.

(1)

- (c) Use the regression equations to compare the relationships between bone length and height for the different bones.
Include in your answer comparisons between male and female ancient Egyptians.

eg As all bone lengths increase the height increases.

eg The female height is estimated to increase more than males for every additional cm of length, with femurs, tibias and humerus.

eg However for the radius the male height is estimated to increase more than females for every additional cm of length.

(3)

Dina suggests comparing the y -intercepts for the regression equations.

- (d) Is Dina's suggestion appropriate?

eg This is not appropriate as the y intercept would relate to a bone length of 0 cm which is unrealistic.

(2)

Higher Paper 2 Q10 a and b

- 10 Researchers used the Petersen capture recapture method to estimate the number of humpback whales off the coasts of California, Oregon and Washington.

They used aerial photographs to identify the whales, recording the whales' unique markings in the first sampling rather than tagging the whales.

In 1995 they recorded the markings of 331 individual whales.

In 1996 they identified 264 individual whales of which 104 had been identified in the previous year.

(Source: *digitalcommons.unl.edu*)

- (a) Calculate an estimate for the number of humpback whales off the coasts of California, Oregon and Washington in 1996

$$\frac{331}{?} = \frac{104}{264}$$

$$? = \frac{331 \times 264}{104} = 840.230..$$

840

(2)

- (b) Discuss the validity and reliability of this estimate.

eg. It is only valid if there have been no births or deaths,
and the markings can be clearly seen.
However the sample sizes are large so would be
reliable.

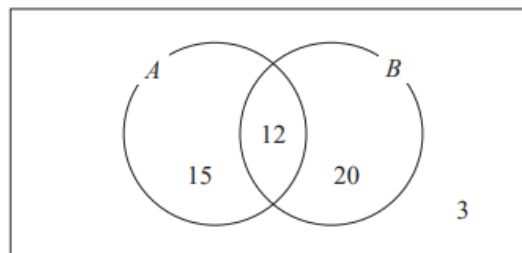
Higher Paper 1 Q11 a and c

11 The Venn diagram shows information about 50 books on a reading list.

A is the event that the book is a science fiction book.

B is the event that the book is an audio book.

The numbers in the Venn diagram show the number of books.



One of the books is chosen at random.

(a) Find $P(A)$.

$$\frac{27}{50}$$

(1)

(b) Find $P(A|B)$.

$$\frac{12}{32}$$

(1)

Mike concludes that A and B are independent events.

(c) Is Mike correct?

Give a reason for your answer.

No, he is wrong. $P(A) \neq P(A|B)$



Examiner Key Notes



Foundation

Based on their performance students should:

- Read each question fully and carefully before attempting to answer it
- Check they have understood the requirement for comments in extended responses and try to give an answer for each section
- Show working out to support the final answer
- Check graphs carefully for any part of the question asking you to complete it. Scatter diagrams will often (but not always) require additional points to be plotted. A bar chart with a blank space in the middle or at the end is likely to mean there is a requirement for them to complete it. Please encourage candidates to check where all marks are being awarded in a question.
- Check the wording of the question too. Often, these values are not required, and we are simply looking for visual comparisons to be made. Usually if a question requires these figures as part of your answer, the question will demand it.
- Appreciate that an item requiring an answer worth two marks will require two parts to their answer. A comment with a conclusion/decision. Or a correlation type with an interpretation.
- Comment on the appropriateness of a conclusion if the question has asked you to. Candidates should not assume we know what their decision is with regards to the appropriateness
 - we need to see a decision. 'No/yes because' is often the minimum required for these types of questions.

Foundation

Practice

- interpreting answers to statistical calculations and diagrams in context
- identifying the appropriateness of calculations, diagrams and approaches
- extended response questions and understand how to break these down with the steps outlined in the question
- key terms and processes as part of the statistical cycle, particularly those exemplified in question 6: hypotheses, pilot studies, improving and writing questionnaires and benefits of sampling
- Candidates should practice on making comparisons, particularly those in context.

Candidates should revise

- how to calculate and interpret relative risk
- how to calculate the proportionate sizes for stratified sampling, and practice describing how to conduct this type of sampling method randomly
- cumulative frequency curves and reading values from it.

Higher

Based on their performance on this paper, students should:

- Show working for statistical calculations
- Ensure that correct statistical language is used throughout when making comparisons, just stating figures, or using the word 'whereas' is not a comparison.

Ensure that they can

- recognise and name different sampling techniques
- describe how to take different types of sampling.

Higher

Practise

- writing clear explanations, bearing in mind exactly what is asked in the question and what evidence you should give to support your answer
- interpreting statistical calculations in the context of the question
- practise giving statistical reasons or against the diagrams suggested
- using the binomial distribution to calculate probabilities
- working with and interpreting standardised scores in a variety of contexts
- calculating relative risk and interpreting relative risk
- analysing plans for data collection and the subsequent diagrams and calculations together with giving statistical reasons or against the approaches suggested
- calculating geometric mean for a range of different contexts.

Develop their understanding of:

- control charts and the actions to be taken for samples falling in different parts of these charts
- independent events
- regression equations and the interpretation of the gradient of these.

Grade boundaries



June 2024

1ST0		9	8	7	6	5	4	3	2	1
Foundation tier						92	70	50	31	12
Higher tier		125	106	87	67	47	28			

(Total marks out of 160)

June 2023

1ST0		9	8	7	6	5	4	3	2	1
Foundation tier						89	71	51	31	11
Higher tier		120	102	84	65	46	27	17		

(Total marks out of 160)

June 2024

Cumulative Percentage by grade at each tier – June 2024

1ST0		9	8	7	6	5	4	3	2	1
Foundation tier	13653					18.2	44.3	67.8	84.0	95.5
Higher tier	12686	8.3	22.9	41.4	63.6	85.0	97.0	99.2		
Total	26339	4.0	11.0	19.9	30.6	50.4	69.7	82.9	91.3	97.3

June 2023

Cumulative Percentage by grade at each tier – June 2023

1ST0		9	8	7	6	5	4	3	2	1
Foundation tier	9721					14.9	40.0	67.0	85.5	97.1
Higher tier	12000	7.6	20.7	37.7	59.1	80.9	95.4	98.4		
Total	21721	4.2	11.4	20.8	32.6	51.4	70.6	84.4	92.7	97.9

Support

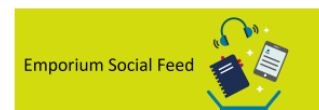
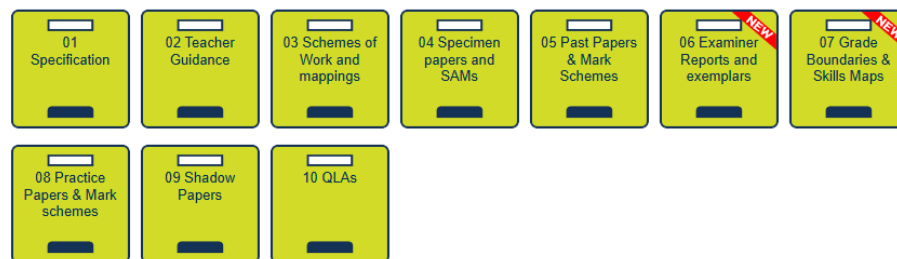


Maths Emporium

Do read the examiners reports.

Category: GCSE Statistics

GCSE Statistics documents for the current 9-1 specification (1ST0)



UPCOMING EVENTS

OCT All day

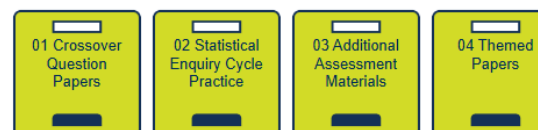
7

**GCSE Statistics Post-Series
Support Summer 2024 Multiple
Releases**

OCT All day

[Maths Emporium](#) > [GCSE Statistics](#) > 08 Practice Papers & Mark schemes

Category: 08 Practice Papers & Mark schemes

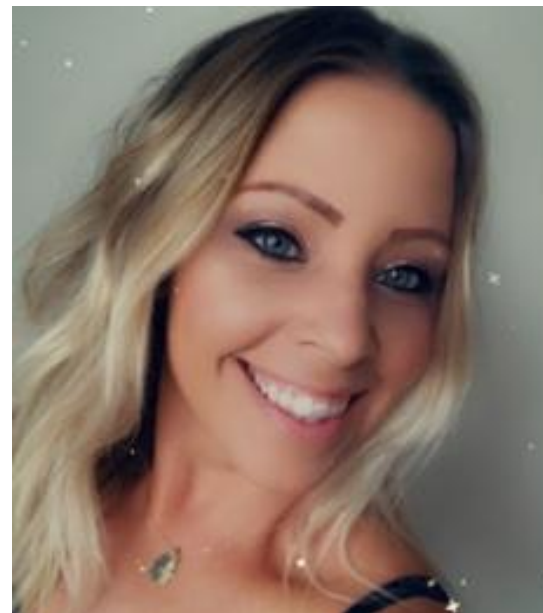


Meet your Maths and Statistics Subject Advisor and Partner



Vicky Wood

Vicky works closely with the wider maths team to support teachers in UK centres in delivering Pearson Edexcel qualifications in Mathematics and Statistics teachingmaths@pearson.com
[Sign up for Vicky's monthly updates](#)



Nicola Woodford-Smith

Nicola works as the maths Subject Partner in the maths team at Pearson Edexcel. She helps to create resources and delivers CPD to support you and your team through the lifecycle of our qualifications.

Follow [@miss_mathsgeek](#) (on X) for updates and information

NEW Podcast: The Right Angle



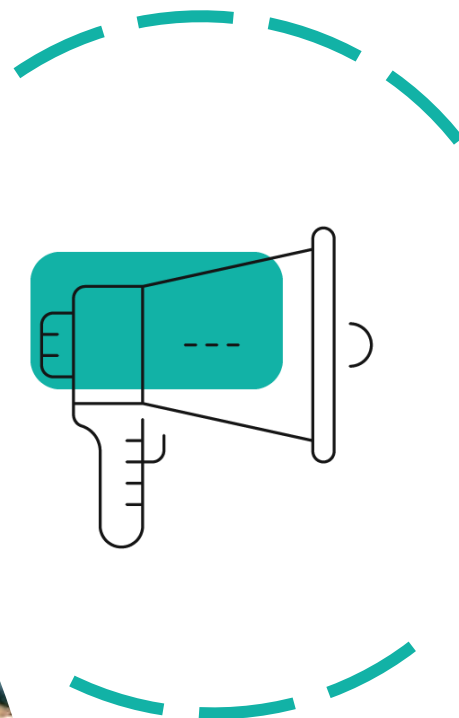
The Right Angle invites topical discussions, debates and insights from a range of thought leaders, award-winning maths educators and facilitators. Our subject partner hosts, Mark Heslop and Nicola Woodford-Smith lead conversations on themes such as the evolution of technology to support learning, student engagement and diversity and inclusion across the education of mathematics. Listen and subscribe for FREE on Apple Podcasts, Spotify and on Soundcloud.



Find out more

For more professional development courses please see Pearson's [Professional Development Academy](#)





Your Feedback Matters

Following this event, you will receive an invitation to share your thoughts about the session. Your feedback is invaluable to us, as it helps us tailor our professional development materials to better meet your needs. Please don't hesitate to let us know what you'd like to see more of and what areas you think could be improved.



Pearson