OCR
Oxford Cambridge and RSA

Friday 14 June 2024 – Afternoon

A Level Further Mathematics B (MEI)

Y422/01 Statistics Major

Time allowed: 2 hours 15 minutes

141469583 6

You must have:
- the Printed Answer Booklet
- the Formulae Booklet for Further Mathematics B (MEI)
- a scientific or graphical calculator

QP

INSTRUCTIONS
- Use black ink. You can use an HB pencil, but only for graphs and diagrams.
- Write your answer to each question in the space provided in the **Printed Answer Booklet**. If you need extra space use the lined page at the end of the Printed Answer Booklet. The question numbers must be clearly shown.
- Fill in the boxes on the front of the Printed Answer Booklet.
- Answer **all** the questions.
- Where appropriate, your answer should be supported with working. Marks might be given for using a correct method, even if your answer is wrong.
- Give your final answers to a degree of accuracy that is appropriate to the context.
- Do **not** send this Question Paper for marking. Keep it in the centre or recycle it.

INFORMATION
- The total mark for this paper is **120**.
- The marks for each question are shown in brackets **[ ]**.
- This document has **12** pages.

ADVICE
- Read each question carefully before you start your answer.

© OCR 2024 [H/508/5594]
DC (PQ/CT) 334980/2

OCR is an exempt Charity

stemup.co.uk

Section A (27 marks)

1 The number of insurance policy sales made per month by a salesperson is modelled by the random variable X, with probability distribution shown in the table.

r	0	1	2	3	4	5	6
$P(X = r)$	0.05	0.1	0.25	0.3	0.15	0.1	0.05

(a) Find each of the following.

- $E(X)$
- $Var(X)$ **[2]**

The salesperson is paid a basic salary of £1000 per month plus £500 for each policy that is sold.

(b) Find the mean and standard deviation of the salesperson's monthly salary. **[3]**

2 The number of cars arriving per minute to queue at a drive-through fast-food restaurant is modelled by the random variable X. The standard deviation of X is 0.6. You should assume that arrivals are random and independent and occur at a constant average rate.

(a) Find the mean of X. **[2]**

(b) (i) Calculate $P(X = 1)$. **[1]**

(ii) Calculate $P(X > 1)$. **[2]**

(c) Find the probability that fewer than 5 cars arrive in a randomly chosen 20-minute period. **[2]**

© OCR 2024

stemup.co.uk

3 At a launderette the process of cleaning a load of clothes consists of three stages: washing, drying and folding. The times in minutes for each process are modelled by independent Normal distributions with means and standard deviations as shown in the table.

	Mean	Standard deviation
Washing	35	2.4
Drying	46	3.1
Folding	12	2.2

(a) Find the probability that drying a randomly chosen load of clothes takes more than 50 minutes. **[1]**

(b) It is given that for 99% of loads of clothes the washing time is less than k minutes.

Find the value of k. **[1]**

(c) Determine the probability that the drying time for a randomly chosen load of clothes is less than the total of the washing and folding times. **[3]**

(d) Determine the probability that the mean time for cleaning 5 randomly chosen loads of clothes is less than 90 minutes. You should assume that the time for cleaning any load is independent of the time for cleaning any other load. **[3]**

4 An archer fires arrows at a circular target of radius 50 cm. The distance in cm that an arrow lands from the centre of the target is modelled by the random variable X, with probability density function given by

$$f(x) = \begin{cases} ax & 0 \leqslant x \leqslant 50, \\ 0 & \text{otherwise,} \end{cases}$$

where a is a constant.

(a) Determine the value of a. **[2]**

(b) Determine the probability that an arrow will land within 5 cm of the centre of the target. **[2]**

(c) Determine the median distance from the centre of the target that an arrow will land. **[3]**

© OCR 2024

stemup.co.uk

Section B (93 marks)

5 A researcher is investigating whether doing yoga has any effect on quality of sleep in older people. The researcher selects a random sample of 40 older people, who then complete a yoga course. Before they start the course and again at the end, the 40 people fill in a questionnaire which measures their perceived sleep quality. The higher the score, the better is the perceived quality of sleep.

The researcher uses software to produce a 90% confidence interval for the difference in mean sleep quality (sleep quality after the course minus sleep quality before the course). The output from the software is shown below.

Z Estimate of a Mean ▼

Confidence level 0.9

Sample

Mean 0.586

s 2.14

N 40

Result

Z Estimate of a Mean

Mean	0.586
s	2.14
SE	0.3384
N	40
Lower limit	0.029
Upper limit	1.143
Interval	0.586 ± 0.557

(a) Explain why the confidence interval is based on the Normal distribution even though the distribution of the population of differences is not known. **[2]**

(b) Explain whether the confidence interval suggests that the mean sleep qualities before and after completing a yoga course are different. **[2]**

(c) In the output from the software, SE stands for 'standard error'.

(i) Explain what standard error is. **[1]**

(ii) Show how the standard error was calculated in this case. **[1]**

(d) A colleague of the researcher suggests that the confidence level should have been 95% rather than 90%.

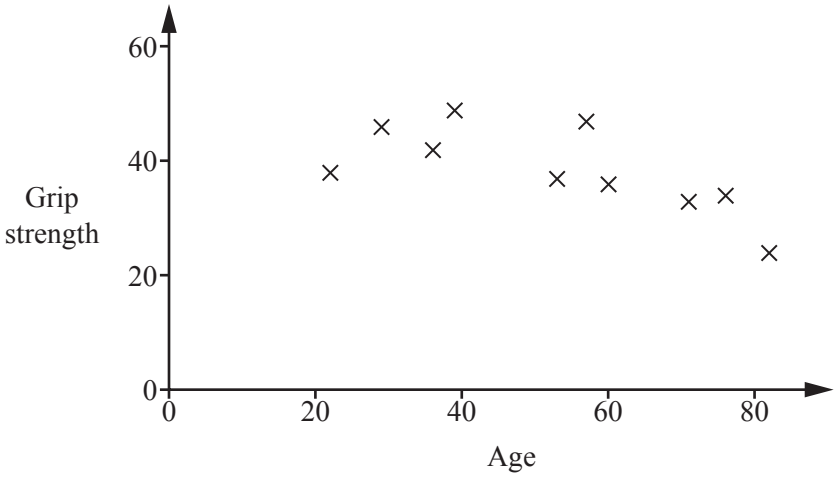
Determine whether this would have made a difference to your answer to part **(b)**. **[4]**

© OCR 2024

stemup.co.uk

6 A student is investigating the relationship between age and grip strength in adults. The student selects 10 people and records their ages in years and the grip strengths of their dominant hand, measured in kg. The data are shown in the table below, together with a scatter diagram to illustrate the data.

Age	22	29	36	39	53	57	60	71	76	82
Grip strength	38	46	42	49	37	47	36	33	34	24



The student decides to carry out a hypothesis test to investigate whether there is negative association between age and grip strength.

(a) Explain why the student decides to carry out a test based on Spearman's rank correlation coefficient. **[2]**

(b) State what property of the sample is required in order for it to be valid to carry out a hypothesis test. **[1]**

(c) **In this question you must show detailed reasoning.**

Assuming that the property in part **(b)** holds, carry out the test at the 5% significance level. **[8]**

stemup.co.uk

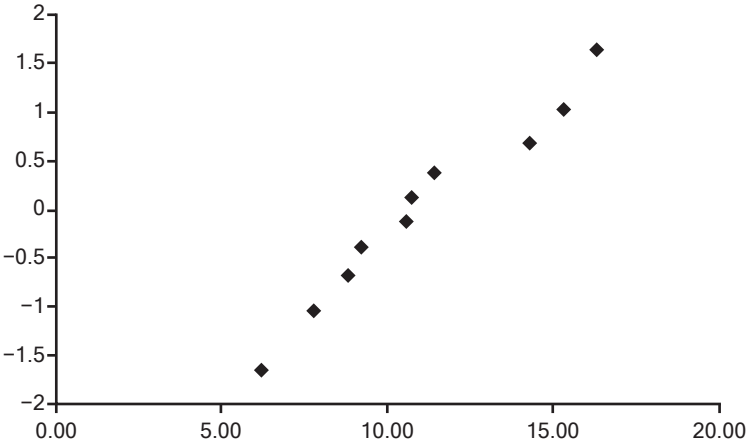
7 An environmental investigator wants to check whether the level of selenium in carrots in fields near a mine is different from the usual level in the country, which is 9.4 ng / g (nanograms per gram). She takes a random sample of 10 carrots from fields near the mine and measures the selenium level of each of them in ng / g, with results as follows.

6.20 10.72 11.42 16.32 15.33 10.56 8.83 9.21 7.78 14.32

(a) Find estimates of each of the following.

- The population mean
- The population standard deviation **[2]**

The investigator produces a Normal probability plot and carries out a Kolmogorov-Smirnov test for these data as shown in the diagram.



Kolmogorov-Smirnov test for Normality

p-value = 0.68

(b) Comment on what the Normal probability plot and the p-value of the test suggest about the data. **[3]**

(c) State the null hypothesis for the Kolmogorov-Smirnov test for Normality. **[1]**

(d) In this question you must show detailed reasoning.

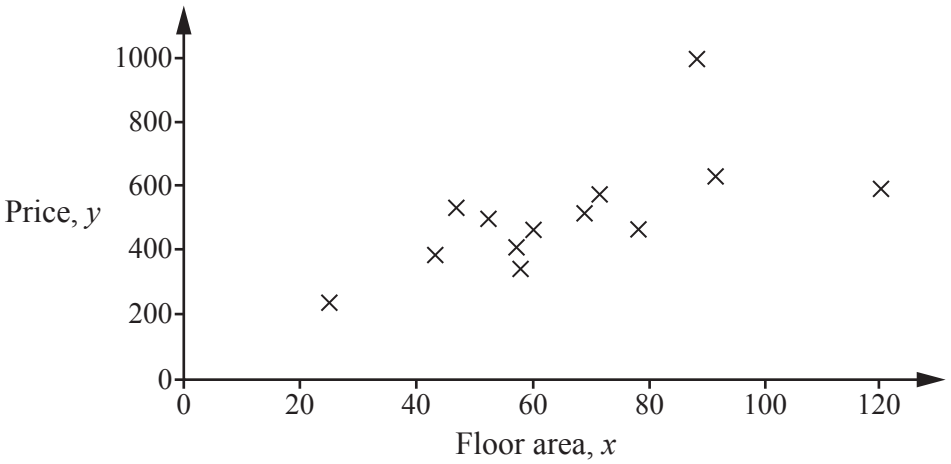
Carry out a test at the 5% significance level to investigate whether the mean selenium level in carrots from fields near the mine is different from 9.4 ng / g. **[8]**

(e) If the p-value of the Kolmogorov-Smirnov test for Normality had been 0.007, explain what procedure you could have used to investigate the selenium level in carrots from fields near the mine. **[2]**

© OCR 2024

stemup.co.uk

8 An estate agent collects data for a random selection of 13 flats in order to investigate the link between the floor areas of flats and their price. The scatter diagram shows the floor areas, $x\,\text{m}^2$, and prices, £y thousand, of the 13 flats.



(a) The estate agent notes that two of the data points are outliers. One is Flat A which has a large floor area but is in poor condition. The other is Flat B which has a balcony with a desirable view overlooking the sea.

Label these two data points on the copy of the scatter diagram in the **Printed Answer Booklet**. **[2]**

The estate agent decides to remove these two data points from the analysis. Summary statistics for the remaining 11 flats are as follows.

$\Sigma x = 652.5 \quad \Sigma y = 5067 \quad \Sigma x^2 = 41\,987.35 \quad \Sigma y^2 = 2\,456\,813 \quad \Sigma xy = 315\,928.2$

(b) **In this question you must show detailed reasoning**.

Calculate the equation of a regression line which is suitable for estimating the price of a flat from its floor area. **[5]**

(c) Use the regression line to estimate the price for the following floor areas.

- $40\,\text{m}^2$
- $110\,\text{m}^2$

[2]

(d) Given that the value of the product moment correlation coefficient for these 11 data items is 0.765, comment on the reliability of your estimates. **[3]**

(e) The estate agent thinks that he can predict the floor area of a flat from its price, using the equation of the regression line found in part **(b)**.

Comment briefly on the estate agent's idea. **[2]**

stemup.co.uk

9 A cyclist has 3 bicycles, a road bike, a gravel bike and an electric bike. She wishes to know if the bicycle which she is riding makes any difference to whether she reaches a speed of 25 mph or greater on a journey. She selects a random sample of 120 journeys and notes the bicycle and whether or not her maximum speed was 25 mph or greater. She decides to carry out a chi-squared test to investigate whether there is any association between bicycle type and whether her maximum speed is 25 mph or greater. **Tables 9.1** and **9.2** show the data and some of the expected frequencies for the test.

Table 9.1

		Bicycle			
		Road	Gravel	Electric	Total
Maximum speed	Less than 25 mph	2	21	19	42
	25 mph or greater	13	47	18	78
	Total	15	68	37	120

Table 9.2

Expected frequency		Bicycle		
		Road	Gravel	Electric
Maximum speed	Less than 25 mph			12.95
	25 mph or greater			24.05

(a) Complete the table of expected frequencies in the **Printed Answer Booklet**. **[2]**

(b) Determine the contribution to the chi-squared test statistic for the Electric bicycle and maximum speed 25 mph or greater. Give your answer correct to **4** decimal places. **[2]**

The contributions to the chi-squared test statistic for the remaining categories are shown in **Table 9.3**.

Table 9.3

Contribution to the test statistic		Bicycle		
		Road	Gravel	Electric
Maximum speed	Less than 25 mph	2.0119	0.3294	2.8264
	25 mph or greater	1.0833	0.1774	

(c) In this question you must show detailed reasoning.

Carry out the test at the 5% significance level. **[6]**

(d) For each type of bicycle, give a brief interpretation of what the data suggest about maximum speed. **[3]**

© OCR 2024

stemup.co.uk

10 Ben takes an underground train to work and back home each day. The waiting time is defined as the time from when he reaches the station platform until he boards the train.

On his way to work the waiting time is X minutes, where X is modelled by a continuous uniform distribution on [0, 6].

On his way back from work, the waiting time is Y minutes, where Y is modelled by a continuous uniform distribution on [0, 4].

Ben's total waiting time for both journeys is Z minutes, where $Z = X + Y$. You should assume that X and Y are independent.

(a) Find E(Z). **[2]**

(b) Ben thinks that Z will be well modelled by a continuous uniform distribution on [0, 10].

By considering variances, show that he is not correct. **[3]**

(c) Ben's friend Jamila constructs the spreadsheet below, which shows a simulation of 20 values of X, Y and Z. All of the values have been rounded to 2 decimal places.

	A	B	C
1	***X***	***Y***	***Z***
2	1.17	3.83	5.01
3	2.01	0.81	2.82
4	1.27	1.52	2.78
5	1.41	3.94	5.35
6	4.11	2.94	7.05
7	1.76	0.96	2.72
8	3.29	0.98	4.27
9	0.77	0.22	0.99
10	0.99	1.44	2.43
11	4.79	2.43	7.22
12	3.82	3.93	7.75
13	5.25	2.74	7.99
14	2.64	0.48	3.12
15	1.54	2.18	3.72
16	2.71	1.66	4.36
17	0.04	3.24	3.28
18	5.95	3.12	9.07
19	5.22	1.21	6.42
20	4.16	0.11	4.27
21	1.02	0.99	2.01
22			

Write down an estimate of P($Z > 6$). **[1]**

(d) Use a Normal approximation to determine the probability that Ben's total waiting time when travelling to and from work on 40 days is more than 210 minutes. **[3]**

stemup.co.uk

11 The discrete random variable X has a uniform distribution over the set of all integers between 25 and n inclusive, where n is a positive integer with $n > 25$.

(a) Determine $\mathrm{P}\left(X < \frac{n+25}{2}\right)$ in each of the following cases.

- n is even
- n is odd **[4]**

(b) Determine an expression in terms of n for the variance of the mean of 100 independent values of X. **[2]**

(c) Given that $n = 75$, calculate an estimate of the probability that the mean of 100 independent values of X is less than 48. **[5]**

12 The cumulative distribution function of the continuous random variable X is given by

$$\mathrm{F}(x) = \begin{cases} 0 & x < 20, \\ a(x^2 + bx + c) & 20 \leqslant x \leqslant 30, \\ 1 & x > 30, \end{cases}$$

where a, b and c are constants.

You are given that $\mathrm{P}(X < 25) = \frac{11}{24}$.

(a) Find $\mathrm{P}(X > 27)$. **[7]**

(b) Find the 90th percentile of X. **[2]**

END OF QUESTION PAPER

© OCR 2024

stemup.co.uk

BLANK PAGE

© OCR 2024

stemup.co.uk

OCR
Oxford Cambridge and RSA

Copyright Information

OCR is committed to seeking permission to reproduce all third-party content that it uses in its assessment materials. OCR has attempted to identify and contact all copyright holders whose work is used in this paper. To avoid the issue of disclosure of answer-related information to candidates, all copyright acknowledgements are reproduced in the OCR Copyright Acknowledgements Booklet. This is produced for each series of examinations and is freely available to download from our public website (www.ocr.org.uk) after the live examination series.

If OCR has unwittingly failed to correctly acknowledge or clear any third-party content in this assessment material, OCR will be happy to correct its mistake at the earliest possible opportunity.

For queries or further information please contact The OCR Copyright Team, The Triangle Building, Shaftesbury Road, Cambridge CB2 8EA.

OCR is part of Cambridge University Press & Assessment, which is itself a department of the University of Cambridge.

© OCR 2024

stemup.co.uk

Surname	StemUp Model Solution
Answer booklet	Write the question number in the left-hand margin at the start of each answer.

Question number

1 (a)

$E(X) = 0 \times 0.05 + 1 \times 0.1 + 2 \times 0.25 + 3 \times 0.3 + 4 \times 0.15 + 5 \times 0.1 + 6 \times 0.05$

$E(X) = 2.9$

$E(X^2) = 0.1 + 1 + 2.7 + 2.4 + 2.5 + 1.8 = 10.5$

$Var(X) = 10.5 - 2.9^2 = 2.09$

1 (b)

Salary $S = 1000 + 500X$

Mean $= 1000 + 500 \times 2.9 = £2450$

$Var(S) = 500^2 \times Var(X) = 500^2 \times 2.09 = 522500$

Standard deviation $= \sqrt{522500} = £723$ (nearest £)

2 (a)

$X \sim Po(\lambda)$, so $Var(X) = mean = \lambda$

$Var(X) = 0.6^2 = 0.36$

Mean of $X = 0.36$

2 (b)(i)

$X \sim Po(0.36)$

$P(X = 1) = e^{-0.36} \times 0.36 = 0.2512$ (4 d.p.)

2 (b)(ii)

$P(X > 1) = 1 - P(X \leq 1)$

$= 1 - 0.9488$

$= 0.0512$ (4 d.p.)

2 (c)

Mean for 20 minutes $= 20 \times 0.36 = 7.2$, $Y \sim Po(7.2)$

$P(Y < 5) = P(Y \leq 4) = 0.1555$ (4 d.p.)

3 (a)

$D \sim N(46, 3.1^2)$

Question number	
	$P(D > 50) = 0.0985$ (3 s.f.)
3 (b)	$W \sim N(35, 2.4^2)$, $P(W < k) = 0.99$
	$k = 35 + 2.3263 \times 2.4 = 40.6$ minutes (3 s.f.)
3 (c)	Let $V = D - (W + F)$
	$E(V) = 46 - 35 - 12 = -1$
	$Var(V) = 3.1^2 + 2.4^2 + 2.2^2 = 9.61 + 5.76 + 4.84 = 20.21$
	$V \sim N(-1, 20.21)$
	$P(D < W + F) = P(V < 0) = 0.588$ (3 s.f.)
3 (d)	Time for one load $T = W + D + F$
	$T \sim N(35 + 46 + 12, 20.21) = N(93, 20.21)$
	Mean of 5 loads: $\bar{T} \sim N(93, \frac{20.21}{5}) = N(93, 4.042)$
	$P(\bar{T} < 90) = 0.0678$ (3 s.f.)
4 (a)	$\int_0^{50} ax\,dx = 1$
	$a[\frac{x^2}{2}]_0^{50} = 1250a = 1$
	$a = \frac{1}{1250}$
4 (b)	$P(X \leq 5) = \int_0^5 \frac{x}{1250}\,dx = \frac{5^2}{2500}$
	$= \frac{1}{100} = 0.01$

Question number	
4 (c)	Median m: $\int_0^m \frac{x}{1250}\,dx = \frac{1}{2}$
	$\frac{m^2}{2500} = \frac{1}{2}$
	$m^2 = 1250$
	$m = \sqrt{1250} = 25\sqrt{2} = 35.4$ cm (3 s.f.)
5 (a)	The sample is large ($n = 40$).
	By the Central Limit Theorem, for a large sample the sample mean
	is approximately Normally distributed, whatever the distribution
	of the population.
5 (b)	The confidence interval (0.029, 1.143) does not contain zero,
	which suggests that there is a difference in the mean sleep
	quality before and after completing the yoga course.
5 (c)(i)	The standard error is the standard deviation of the sample mean.
5 (c)(ii)	$SE = \frac{s}{\sqrt{n}} = \frac{2.14}{\sqrt{40}} = 0.3384$
5 (d)	95% interval: $0.586 \pm 1.96 \times 0.3384$
	$= 0.586 \pm 0.6633$
	Interval is $(-0.077, 1.249)$
	This interval does contain zero, so it suggests that there may be
	no difference in the mean sleep quality.
	So yes, using 95% would have changed the answer to part (b).

Question number

6 (a) The scatter diagram looks curvilinear rather than elliptical, so there is not enough evidence to suggest that the underlying distribution is bivariate Normal (so a PMCC test is not suitable).

6 (b) The sample must be a random sample.

6 (c)

Age	22	29	36	39	53	57	60	71	76	82
Grip	38	46	42	49	37	47	36	33	34	24
Rank A	1	2	3	4	5	6	7	8	9	10
Rank G	6	8	7	10	5	9	4	2	3	1
d	−5	−6	−4	−6	0	−3	3	6	6	9
d^2	25	36	16	36	0	9	9	36	36	81

$\Sigma d^2 = 284$

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)} = 1 - \frac{6 \times 284}{10 \times 99} = -0.7212 \text{ (4 d.p.)}$$

H_0: there is no association between age and grip strength in the population

H_1: there is negative association between age and grip strength in the population

$n = 10$, 5% one-tailed critical value = 0.5636

$-0.7212 < -0.5636$, so the result is significant: reject H_0.

There is sufficient evidence at the 5% level to suggest that there is negative association between age and grip strength.

7 (a) $n = 10$, $\Sigma x = 110.69$, $\Sigma x^2 = 1325.0235$

Estimate of population mean $\bar{x} = 11.069 = 11.07$ (4 s.f.)

Question number

$s^2 = \frac{1}{9}(1325.0235 - \frac{110.69^2}{10}) = 11.088$

Estimate of population standard deviation $s = 3.33$ (3 s.f.)

7 (b) The Normal probability plot is roughly a straight line.

The p-value of 0.68 is very high (well above 0.05).

Both suggest that the data may come from a Normal distribution.

7 (c) H_0: the Normal distribution is a good fit to the data

(the data come from a Normally distributed population).

7 (d) H_0: $\mu = 9.4$, H_1: $\mu \neq 9.4$

where μ is the population mean selenium level (ng/g) in carrots

from fields near the mine.

Population assumed Normal, σ unknown, so use a t test with 9 d.f.

Test statistic $t = \frac{11.069 - 9.4}{3.3299 / \sqrt{10}} = 1.585$ (4 s.f.)

Refer to t_9: 5% two-tailed critical value = 2.262

1.585 < 2.262, so the result is not significant: do not reject H_0.

There is insufficient evidence at the 5% level to suggest that the

mean selenium level in carrots near the mine is different from

9.4 ng/g.

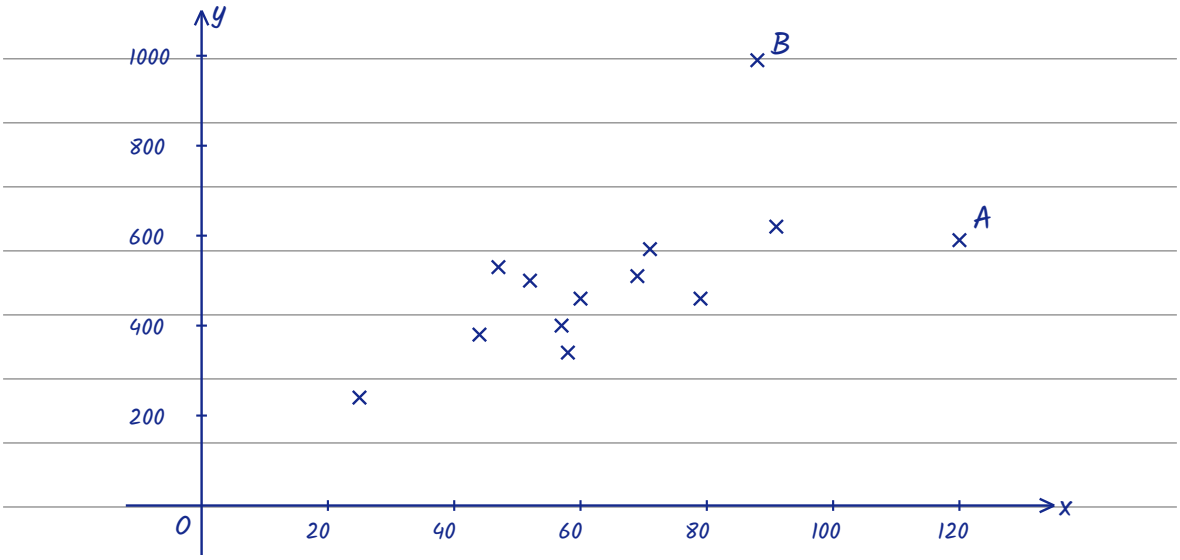
7 (e) A p-value of 0.007 would suggest that the population is unlikely

to be Normally distributed, so the t test would not be valid.

Instead use a Wilcoxon signed rank test (on the median).

Question number

8 (a)



A = (120, 590): large floor area but low price (poor condition)

B = (88, 990): very high price for its area (sea view balcony)

8 (b)

Price depends on floor area, so use the regression line of y on x.

$\bar{x} = \frac{652.5}{11} = 59.318, \quad \bar{y} = \frac{5067}{11} = 460.64$

$S_{xx} = 41987.35 - \frac{652.5^2}{11} = 3282.24$

$S_{xy} = 315928.2 - \frac{652.5 \times 5067}{11} = 15362.97$

$b = \frac{S_{xy}}{S_{xx}} = \frac{15362.97}{3282.24} = 4.6806$

$y - 460.64 = 4.6806(x - 59.318)$

$y = 4.681x + 183.0$

8 (c)

$x = 40$: $y = 4.6806 \times 40 + 182.99 = 370.2 \Rightarrow$ £370 thousand

$x = 110$: $y = 4.6806 \times 110 + 182.99 = 697.9 \Rightarrow$ £698 thousand

8 (d)

$r^2 = 0.765^2 = 0.585$, which is not very close to 1, so the points

Question number

are not that close to the line: estimates are only moderately reliable.

40 m^2 lies within the range of the data (interpolation), so this estimate is fairly reliable, but not very, as the points are scattered.

110 m^2 is outside the range of the data (extrapolation), so this estimate is even less reliable.

8 (e) This is not sensible. The line in part (b) is the regression line of y on x, which only estimates the price for a given floor area. To predict floor area from price he would need the regression line of x on y (floor area on price).

9 (a)

Expected	Road	Gravel	Electric
< 25 mph	5.25	23.80	12.95
≥ 25 mph	9.75	44.20	24.05

eg Road, < 25: $\frac{42 \times 15}{120} = 5.25$, Gravel, < 25: $\frac{42 \times 68}{120} = 23.80$

9 (b) $\frac{(O-E)^2}{E} = \frac{(18-24.05)^2}{24.05} = 1.5219$ (4 d.p.)

9 (c) H_0: there is no association between type of bicycle and whether maximum speed is 25 mph or greater

H_1: there is some association between type of bicycle and whether maximum speed is 25 mph or greater

$X^2 = 2.0119 + 0.3294 + 2.8264 + 1.0833 + 0.1774 + 1.5219$

$X^2 = 7.9503$

Degrees of freedom $= (2-1)(3-1) = 2$

5% critical value of $\chi^2_2 = 5.991$

Question number	
	7.9503 > 5.991, so the result is significant: reject H_0.
	There is sufficient evidence at the 5% level to suggest that there is some association between type of bicycle and whether maximum speed is 25 mph or greater.
9 (d)	Road: contributions 2.0119 and 1.0833 show fewer journeys than expected with max speed < 25 mph and more than expected ≥ 25 mph.
	Gravel: contributions 0.3294 and 0.1774 are small, so the numbers are about as expected.
	Electric: contributions 2.8264 and 1.5219 show more journeys than expected with max speed < 25 mph and fewer than expected ≥ 25 mph.
10 (a)	$E(X) = \frac{0+6}{2} = 3, \quad E(Y) = \frac{0+4}{2} = 2$
	$E(Z) = 3 + 2 = 5$
10 (b)	$Var(X) = \frac{6^2}{12} = 3, \quad Var(Y) = \frac{4^2}{12} = \frac{4}{3}$
	$Var(Z) = 3 + \frac{4}{3} = \frac{13}{3} = 4.33$ (X, Y independent)
	Uniform on [0, 10] would have variance $\frac{10^2}{12} = \frac{25}{3} = 8.33$
	$\frac{13}{3}$ is very different from $\frac{25}{3}$, so Ben is not correct.
10 (c)	Values of Z above 6: 7.05, 7.22, 7.75, 7.99, 9.07, 6.42 (6 of 20)
	$P(Z > 6) \approx \frac{6}{20} = 0.3$

Question number

10 (d) Total T = sum of 40 independent values of Z

$E(T) = 40 \times 5 = 200, \quad Var(T) = 40 \times \frac{13}{3} = \frac{520}{3} = 173.33$

By the Central Limit Theorem, $T \approx N(200, 173.33)$

$P(T > 210) = P(Z > \frac{210 - 200}{\sqrt{173.33}}) = P(Z > 0.7596)$

$= 0.224$ (3 s.f.)

11 (a) X takes the values 25, 26, ..., n: there are $n - 24$ values,

each with probability $\frac{1}{n - 24}$.

n even: $\frac{n + 25}{2}$ is not an integer; values below it are

25, ..., $\frac{n + 24}{2}$, that is $\frac{n + 24}{2} - 24 = \frac{n - 24}{2}$ values

$P(X < \frac{n + 25}{2}) = \frac{n - 24}{2} \times \frac{1}{n - 24} = \frac{1}{2}$

n odd: $\frac{n + 25}{2}$ is an integer; values below it are

25, ..., $\frac{n + 23}{2}$, that is $\frac{n + 23}{2} - 24 = \frac{n - 25}{2}$ values

$P(X < \frac{n + 25}{2}) = \frac{n - 25}{2(n - 24)}$

11 (b) Uniform on $n - 24$ consecutive integers:

$Var(X) = \frac{1}{12}((n - 24)^2 - 1)$

Question number

$\text{Var}(\bar{X}) = \frac{\text{Var}(X)}{100} = \frac{(n-24)^2 - 1}{1200} = \frac{n^2 - 48n + 575}{1200}$

11 (c) $n = 75$: $E(X) = \frac{25 + 75}{2} = 50$

$\text{Var}(\bar{X}) = \frac{51^2 - 1}{1200} = \frac{2600}{1200} = \frac{13}{6} = 2.1667$

By the Central Limit Theorem, $\bar{X} \approx N(50, \frac{13}{6})$

Total of 100 values is an integer, so $\bar{X}$ takes values in steps of 0.01:

continuity correction $P(\bar{X} < 48) = P(\bar{X} < 47.995)$

$= P(Z < \frac{47.995 - 50}{\sqrt{13/6}}) = P(Z < -1.3621)$

$= 0.0866$ (3 s.f.)

12 (a) $F(20) = 0$: $a(400 + 20b + c) = 0$ ...(1)

$F(30) = 1$: $a(900 + 30b + c) = 1$ ...(2)

$F(25) = \frac{11}{24}$: $a(625 + 25b + c) = \frac{11}{24}$ ...(3)

(2) − (1): $a(500 + 10b) = 1$

(3) − (1): $a(225 + 5b) = \frac{11}{24} \Rightarrow a(450 + 10b) = \frac{11}{12}$

Subtract: $50a = 1 - \frac{11}{12} = \frac{1}{12} \Rightarrow a = \frac{1}{600}$

$500 + 10b = 600 \Rightarrow b = 10$

$400 + 200 + c = 0 \Rightarrow c = -600$

$P(X > 27) = 1 - F(27) = 1 - \frac{1}{600}(27^2 + 10 \times 27 - 600)$

Question number

$= 1 - \frac{399}{600} = \frac{67}{200} = 0.335$

12 (b) $\frac{1}{600}(x^2 + 10x - 600) = 0.9$

$x^2 + 10x - 1140 = 0$

$x = \frac{-10 + \sqrt{100 + 4560}}{2}$ (positive root, $20 \leq x \leq 30$)

90th percentile = 29.1 (3 s.f.)