OCR
Oxford Cambridge and RSA

Friday 17 June 2022 – Afternoon

A Level Further Mathematics B (MEI)

Y422/01 Statistics Major

Time allowed: 2 hours 15 minutes

9452143818

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

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 pages 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 2022 [H/508/5594]
DC (ST/SG) 300809/4
OCR is an exempt Charity

stemup.co.uk

Answer **all** the questions.

Section A (25 marks)

1 During a meteor shower, the number of meteors that can be seen at a particular location can be modelled by a Poisson distribution with mean 1.2 per minute.

(a) Find the probability that exactly 2 meteors are seen in a period of 1 minute. **[1]**

(b) Find the probability that more than 3 meteors are seen in a period of 1 minute. **[2]**

(c) Find the probability that no more than 8 meteors are seen in a period of 10 minutes. **[2]**

(d) Explain what the fact that the number of meteors seen can be modelled by a Poisson distribution tells you about the occurrence of meteors. **[2]**

2 A manufacturer is testing how long coloured LED lights will last before the battery runs out, using two different battery types. The times in hours before the battery runs out are modelled by independent Normal distributions with means and standard deviations as shown in the table.

	Time	
Type	Mean	Standard deviation
A	23	2.8
B	35	3.6

(a) In a particular test, a battery of type A is used and the time taken for it to run out is recorded. This process is repeated until a total of 5 randomly selected batteries have been used.

Determine the probability that the total time the 5 batteries last is at least 120 hours. **[3]**

(b) In a similar test, 3 randomly selected batteries of type A are used, one after the other. Then 2 randomly selected batteries of type B are used, one after the other.

Determine the probability that the 3 type A batteries last longer in total than the 2 type B batteries. **[3]**

(c) Explain why it is necessary that the Normal distributions are independent in order to be able to find the probability in part **(b)**. **[1]**

© OCR 2022

stemup.co.uk

3 The table shows the probability distribution of the random variable X, where a and b are constants.

r	0	1	2	3	4
$P(X = r)$	a	b	0.24	0.32	b^2

(a) Given that $E(X) = 1.8$, determine the values of a and b. **[3]**

The random variable Y is given by $Y = 10 - 3X$.

(b) Using the values of a and b which you found in part **(a)**, find each of the following.

- $E(Y)$
- $Var(Y)$ **[3]**

4 A pack of k cards is labelled 1, 2, …, k. A card is drawn at random from the pack. The random variable X represents the number on the card.

(a) Given that $k > 10$, find $P(X \geqslant 10)$. **[2]**

You are now given that $k = 20$.

(b) A card is drawn at random from the pack and the number on it is noted. The card is then returned to the pack. This process is repeated until the second occasion on which the number noted is less than 9.

Find the probability that no more than 4 cards have to be drawn. **[3]**

© OCR 2022

stemup.co.uk

Answer **all** the questions.

Section B (95 marks)

5 A motorist is investigating the relationship between tyre pressure and temperature. As the temperature increases during a hot day, she records the pressure (measured in bars) of one of her car tyres at specific temperatures of 20 °C, 22 °C, …, 36 °C. The results are shown in **Table 5.1**.

Temperature (t °C)	20	22	24	26	28	30	32	34	36
Tyre pressure (P bar)	2.012	2.036	2.065	2.074	2.114	2.140	2.149	2.176	2.192

Table 5.1

(a) Calculate the equation of the regression line of pressure on temperature. Give your answer in the form $P = at + b$, giving the values of a and b to **4** significant figures. **[2]**

(b) **Table 5.2** shows the residuals for most of the data values. Complete the copy of the table in the Printed Answer Booklet. **[3]**

Temperature	20	22	24	26	28	30	32	34	36
Residual tyre pressure	−0.003	−0.002	0.004	−0.010		0.011	−0.003	0.001	

Table 5.2

(c) With reference to the values of the residuals, comment on the goodness of fit of the regression line. **[2]**

(d) Use your answer to part **(a)** to calculate an estimate of the pressure in the tyre at each of the following temperatures, giving your answers to **3** decimal places.

- 25 °C
- 10 °C

[2]

(e) Comment on the reliability of each of your estimates. **[2]**

© OCR 2022

stemup.co.uk

6 In this question you must show detailed reasoning.

A particular type of skin cream is meant to contain an average of 12.2 g of liquid paraffin in a 200 g tub. A researcher believes that the average may differ from this. A random sample of 40 tubs is chosen, and the weight of liquid paraffin, x g, in each tub is measured. You are given that

$\sum x = 491.84, \qquad \sum x^2 = 6050.3.$

(a) Determine a 95% confidence interval for the mean weight of liquid paraffin in a tub. **[6]**

(b) Explain whether the confidence interval supports the researcher's belief. **[1]**

(c) Explain why the sample has to be random in order to construct the confidence interval. **[1]**

(d) A 95% confidence interval for the mean weight in grams of another ingredient in the skin cream is [1.202, 1.398]. This confidence interval is based on a large sample and the unbiased estimate of the population variance calculated from the sample is 0.25.

Find each of the following.

- The mean of the sample
- The size of the sample **[3]**

7 Amir is trying to thread a needle. On each attempt the probability that he is successful is 0.3, independently of any other attempt. The random variable X represents the number of attempts that he takes to thread the needle.

(a) Find $P(X = 5)$. **[2]**

(b) During the course of a day, Amir has to thread 6 needles.

Determine the probability that it takes him more than 3 attempts to be successful for at least 4 of the 6 needles. **[3]**

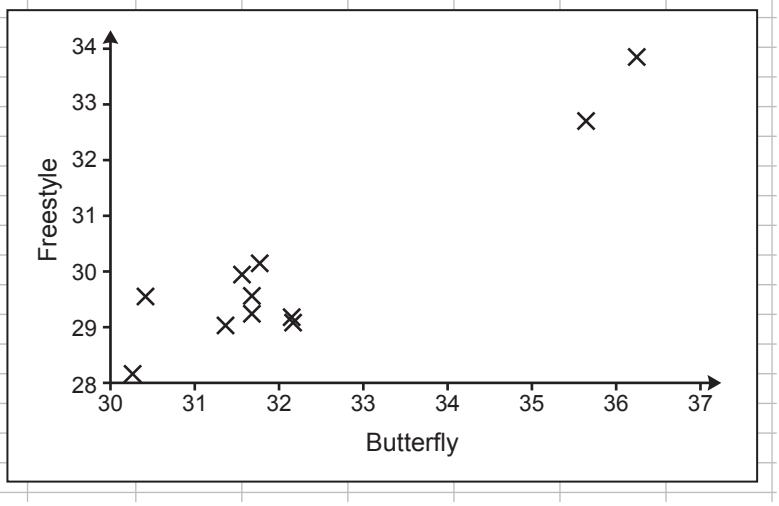
(c) Amaya is also trying to thread a needle. On each attempt the probability that she is successful is p, independently of any other attempt. The probability that Amaya takes 2 attempts to thread a particular needle is $\frac{28}{121}$.

Determine the possible values of p. **[3]**

stemup.co.uk

8 A swimming coach is investigating whether there is correlation between the times taken by teenage swimmers to swim 50 m Butterfly and 50 m Freestyle. The coach selects a random sample of 11 teenage swimmers and records the times that each of them take for each event. The spreadsheet shows the data, together with a scatter diagram to illustrate the data.

	A	B
1	Butterfly	Freestyle
2	30.26	28.19
3	30.41	29.59
4	31.36	29.07
5	31.56	29.99
6	31.68	29.28
7	31.69	29.60
8	31.77	30.18
9	32.14	29.20
10	32.16	29.12
11	35.63	32.69
12	36.24	33.85



(a) In the scatter diagram, Butterfly times have been plotted on the horizontal axis and Freestyle times on the vertical axis. A student states that the variables should have been plotted the other way around.

Explain whether the student is correct. **[2]**

The student decides to carry out a hypothesis test to investigate whether there is any correlation between the times taken for the two events.

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

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

Carry out the test at the 5% significance level. **[8]**

(d) The student concludes that there is definitely no correlation between the times.

Comment on the student's conclusion. **[2]**

© OCR 2022

stemup.co.uk

9 The random variable X has a discrete uniform distribution over the values $\{0, 1, 2, \ldots, 20\}$.

(a) Find $P(X \leqslant 7)$. **[1]**

(b) Find each of the following.

- $E(X)$
- $Var(X)$ **[3]**

The spreadsheet shows a simulation of the distribution of X. Each of the 25 rows of the spreadsheet below the heading row shows a simulation of 10 independent values of X together with the value of the mean of the 10 values, denoted by Y.

	A	B	C	D	E	F	G	H	I	J	K	L
1	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	Y	
2	1	6	2	1	18	6	4	9	11	11	6.9	
3	13	14	12	2	4	11	16	0	16	0	8.8	
4	4	17	1	16	4	10	12	2	18	13	9.7	
5	2	8	12	1	4	16	12	2	15	8	8.0	
6	7	15	16	0	4	7	1	13	0	20	8.3	
7	15	13	10	1	12	0	20	15	16	6	10.8	
8	14	13	17	12	2	18	16	18	9	4	12.3	
9	20	2	12	3	17	3	0	18	15	13	10.3	
10	2	12	5	12	2	6	0	9	10	15	7.3	
11	5	11	13	10	9	17	10	4	20	15	11.4	
12	14	9	9	7	6	20	2	2	11	16	9.6	
13	15	19	18	19	7	6	6	20	3	8	12.1	
14	5	10	6	4	1	19	15	8	17	18	10.3	
15	0	3	15	15	11	12	0	3	9	16	8.4	
16	1	12	1	15	0	4	11	11	9	2	6.6	
17	12	5	0	8	3	8	12	19	13	12	9.2	
18	9	5	1	13	5	4	18	1	1	19	7.6	
19	16	2	20	20	12	17	2	7	8	20	12.4	
20	18	17	3	2	8	18	7	0	11	6	9.0	
21	15	10	7	20	4	0	5	6	11	14	9.2	
22	3	9	10	14	2	1	8	6	0	7	6.0	
23	11	10	11	10	19	11	3	7	10	0	9.2	
24	12	14	6	6	5	20	11	18	10	14	11.6	
25	1	11	5	14	11	10	1	1	2	0	5.6	
26	0	14	7	11	18	5	10	20	11	9	10.5	
27												

(c) Use the spreadsheet to estimate $P(Y \leqslant 7)$. **[1]**

(d) Explain why the true value of $P(Y \leqslant 7)$ is less than $P(X \leqslant 7)$, relating your answer to $Var(X)$ and $Var(Y)$. **[2]**

(e) The random variable W is the mean of 30 independent values of X.

Determine an estimate of $P(W \leqslant 7)$. **[4]**

stemup.co.uk

10 A scientist is researching dietary fat intake and cholesterol level. A random sample of 60 people is selected and their dietary fat intakes and cholesterol levels are measured. Dietary fat intakes are classified as low, medium and high, and cholesterol levels are classified as normal and high.

The scientist decides to carry out a chi-squared test to investigate whether there is any association between dietary fat intake and cholesterol level. **Tables 10.1** and **10.2** show the data and some of the expected frequencies for the test.

		Dietary fat intake			
		Low	**Medium**	**High**	**Total**
Cholesterol level	**Normal**	9	18	5	32
	High	3	13	12	28
	Total	12	31	17	60

Table 10.1

Expected frequency		**Dietary fat intake**		
		Low	**Medium**	**High**
Cholesterol level	**Normal**			9.0667
	High			7.9333

Table 10.2

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

(b) Determine the contribution to the chi-squared test statistic for people with normal cholesterol level and high dietary fat intake, giving your answer to **4** decimal places. **[2]**

The contributions to the chi-squared test statistic for the remaining categories are shown in **Table 10.3**.

Contribution to the test statistic		**Dietary fat intake**		
		Low	**Medium**	**High**
Cholesterol level	**Normal**	1.0563	0.1301	
	High	1.2071	0.1487	2.0846

Table 10.3

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

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

(d) For each level of dietary fat intake, give a brief interpretation of what the data suggest about the level of cholesterol. **[3]**

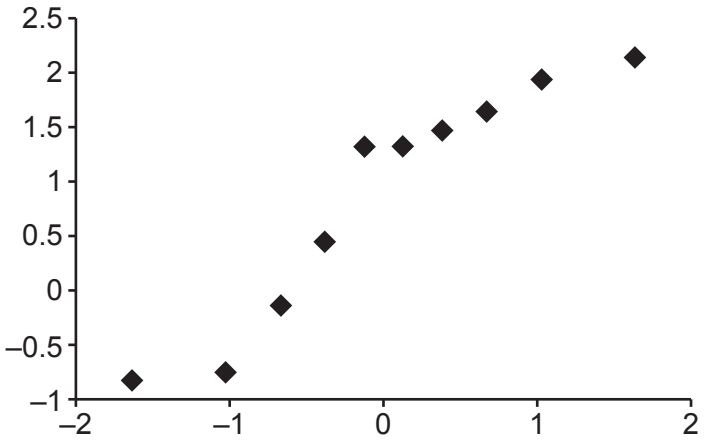
© OCR 2022

stemup.co.uk

11 A particular dietary supplement, when taken for a period of 1 month, is claimed to increase lean body mass of adults by an average of 1 kg. A researcher believes that this claim overestimates the increase. She selects a random sample of 10 adults who then each take the supplement for a month. The increases in lean body masses in kg are as follows.

−0.84 −0.76 −0.16 0.43 1.31 1.32 1.47 1.64 1.93 2.14

A Normal probability plot and the *p*-value of the Kolmogorov-Smirnov test for these data are shown below.



Kolmogorov-Smirnov test for Normality

p-value = 0.049

(a) The researcher decides to carry out a hypothesis test in order to investigate the claim.

Comment on the type of hypothesis test that should be used. You should refer to

- The Normal probability plot
- The *p*-value of the Kolmogorov-Smirnov test **[3]**

(b) Carry out a test at the 5% significance level to investigate whether the researcher's belief may be correct. **[7]**

(c) If the Normal probability plot had been different, giving a *p*-value of 0.65 for the Kolmogorov-Smirnov test, a different procedure could have been used to investigate the researcher's belief.

- State what alternative test could have been used in this case.
- State what the hypotheses would have been. **[3]**

stemup.co.uk

12 The continuous random variable X has cumulative distribution function given by

$$\mathrm{F}(x) = \begin{cases} 0 & x < 0, \\ k(ax - 0.5x^2) & 0 \leqslant x \leqslant a, \\ 1 & x > a, \end{cases}$$

where a and k are positive constants.

(a) Determine the median of X in terms of a. **[7]**

(b) Given that $a = 10$, determine the probability that X is within one standard deviation of its mean. **[7]**

END OF QUESTION PAPER

© OCR 2022

stemup.co.uk

BLANK PAGE

© OCR 2022

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 2022

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) $X \sim Po(1.2)$

$$P(X = 2) = \frac{e^{-1.2} \times 1.2^2}{2!} = 0.2169 \text{ (4 d.p.)}$$

1 (b) $P(X > 3) = 1 - P(X \leq 3)$

$= 1 - 0.9662$

$= 0.0338$ (4 d.p.)

1 (c) Mean for 10 minutes $= 10 \times 1.2 = 12$, $Y \sim Po(12)$

$P(Y \leq 8) = 0.1550$ (4 d.p.)

1 (d) Meteors are seen randomly and independently of each other, and at a uniform (constant) average rate.

2 (a) $T = A_1 + A_2 + A_3 + A_4 + A_5$

$T \sim N(5 \times 23, 5 \times 2.8^2) = N(115, 39.2)$

$$P(T \geq 120) = P\left(Z \geq \frac{120 - 115}{\sqrt{39.2}}\right) = P(Z \geq 0.7986)$$

$= 0.212$ (3 s.f.)

2 (b) $D = (A_1 + A_2 + A_3) - (B_1 + B_2)$

$E(D) = 3 \times 23 - 2 \times 35 = 69 - 70 = -1$

$Var(D) = 3 \times 2.8^2 + 2 \times 3.6^2 = 23.52 + 25.92 = 49.44$

$D \sim N(-1, 49.44)$

Question number

$P(D > 0) = P\left(Z > \frac{0 - (-1)}{\sqrt{49.44}}\right) = P(Z > 0.1422)$

$= 0.443$ (3 s.f.)

2 (c) The variance of a sum or difference is only the sum of the variances if the variables are independent.

3 (a) $E(X) = 0 \times a + 1 \times b + 2 \times 0.24 + 3 \times 0.32 + 4 \times b^2 = 1.8$

$b + 0.48 + 0.96 + 4b^2 = 1.8$

$4b^2 + b - 0.36 = 0$

$b = \frac{-1 \pm \sqrt{1 + 5.76}}{8} = \frac{-1 \pm 2.6}{8}$

$b = 0.2$ ($b = -0.45$ rejected as a probability cannot be negative)

$a + b + 0.24 + 0.32 + b^2 = 1$

$a = 1 - 0.2 - 0.24 - 0.32 - 0.04 = 0.2$

$a = 0.2,\ b = 0.2$

3 (b) $E(Y) = 10 - 3E(X) = 10 - 3 \times 1.8 = 4.6$

$E(X^2) = 1 \times 0.2 + 4 \times 0.24 + 9 \times 0.32 + 16 \times 0.04$

$= 0.2 + 0.96 + 2.88 + 0.64 = 4.68$

$Var(X) = 4.68 - 1.8^2 = 1.44$

$Var(Y) = 3^2 \times Var(X) = 9 \times 1.44 = 12.96$

4 (a) Cards 10, 11, ..., k: there are $k - 9$ of them

$P(X \geq 10) = \frac{k - 9}{k}$

Question number	
4 (b)	$P(\text{number} < 9) = \frac{8}{20} = 0.4$, $P(\text{number} \geqslant 9) = 0.6$
	Need the 2nd success on the 2nd, 3rd or 4th draw:
	$P = 0.4^2 + 2 \times 0.4^2 \times 0.6 + 3 \times 0.4^2 \times 0.6^2$
	$= 0.16 + 0.192 + 0.1728$
	$= 0.5248$
5 (a)	$\Sigma t = 252$, $\Sigma P = 18.958$, $n = 9$
	$S_{tt} = 240$, $S_{tP} = 2.748$
	$a = \frac{S_{tP}}{S_{tt}} = \frac{2.748}{240} = 0.01145$
	$b = \bar{P} - a\bar{t} = 2.10644 - 0.01145 \times 28 = 1.786$
	$P = 0.01145t + 1.786$
5 (b)	Residual = observed − predicted
	$t = 28$: $2.114 - (0.01145 \times 28 + 1.786) = 2.114 - 2.1066 = 0.007$
	$t = 36$: $2.192 - (0.01145 \times 36 + 1.786) = 2.192 - 2.1982 = -0.006$

t	20	22	24	26	28	30	32	34	36
Residual	−0.003	−0.002	0.004	−0.010	0.007	0.011	−0.003	0.001	−0.006

5 (c)	The residuals are all very small, so the line fits the data well.
	There is no pattern in the signs of the residuals, so there is no
	suggestion that the relationship is non-linear.
5 (d)	$t = 25$: $P = 0.01145 \times 25 + 1.786 = 2.072$ bar (3 d.p.)
	$t = 10$: $P = 0.01145 \times 10 + 1.786 = 1.900$ bar (3 d.p.)
5 (e)	25 °C: interpolation and the residuals are small (points close

Question number

to the line), so this estimate is likely to be reliable.

10 °C: this is extrapolation, outside the range 20 to 36 °C,

so this estimate is much less reliable.

6 (a) $\bar{x} = \frac{491.84}{40} = 12.296$

$S_{xx} = 6050.3 - \frac{491.84^2}{40} = 2.63536$

$s^2 = \frac{2.63536}{39} = 0.067573, \quad s = 0.25995$

$n = 40$ is large, so by the CLT $\bar{X}$ is approximately Normal

95% CI: $\bar{x} \pm 1.96 \times \frac{s}{\sqrt{n}}$

$= 12.296 \pm 1.96 \times \frac{0.25995}{\sqrt{40}}$

$= 12.296 \pm 0.081$

(12.215, 12.377)

6 (b) The interval does not contain 12.2, so it supports

the researcher's belief that the average may differ from 12.2 g.

6 (c) A random sample is needed so that valid inferences can be made

about the population from the sample.

6 (d) Sample mean $= \frac{1.202 + 1.398}{2} = 1.3$

Half-width $= \frac{1.398 - 1.202}{2} = 0.098$

Question number	
	$1.96 \times \frac{\sqrt{0.25}}{\sqrt{n}} = 0.098$
	$\sqrt{n} = \frac{1.96 \times 0.5}{0.098} = 10$
	Sample size $n = 100$
7 (a)	$X \sim Geo(0.3)$
	$P(X = 5) = 0.7^4 \times 0.3 = 0.07203$
7 (b)	$P(X > 3) = 0.7^3 = 0.343$
	N = number of needles (out of 6) taking more than 3 attempts
	$N \sim B(6, 0.343)$
	$P(N \geq 4) = 1 - P(N \leq 3) = 1 - 0.8900$
	$= 0.1100$ (4 d.p.)
7 (c)	$P(2 \text{ attempts}) = (1 - p)p = \frac{28}{121}$
	$121p - 121p^2 = 28$
	$121p^2 - 121p + 28 = 0$
	$(11p - 4)(11p - 7) = 0$
	$p = \frac{4}{11}$ or $p = \frac{7}{11}$
8 (a)	The student is not correct. Both variables are random
	(neither is controlled), so it is equally valid to plot
	them either way around.
8 (b)	The scatter diagram is not elliptical (two outliers at the top

Question number

right), so the data probably do not come from a bivariate Normal distribution, so a pmcc test is not appropriate.

8 (c)

H_0: there is no association between Butterfly and Freestyle times in the population of teenage swimmers

H_1: there is some association between Butterfly and Freestyle times in the population of teenage swimmers

Rank B	1	2	3	4	5	6	7	8	9	10	11
Rank F	1	6	2	8	5	7	9	4	3	10	11
d^2	0	16	1	16	0	1	4	16	36	0	0

$\Sigma d^2 = 90$

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2-1)} = 1 - \frac{6 \times 90}{11 \times 120} = 0.5909$$

For $n = 11$, 5% two-tailed critical value = 0.6182

$0.5909 < 0.6182$, so the result is not significant: do not reject H_0.

There is insufficient evidence at the 5% level to suggest that there is association between Butterfly and Freestyle times.

8 (d)

The conclusion is not certain. Not rejecting H_0 does not prove there is no correlation: a different sample, or a different significance level, might give a different conclusion.

9 (a)

$P(X \leq 7) = \frac{8}{21}$ $(= 0.381)$

9 (b)

$E(X) = \frac{0 + 20}{2} = 10$

Question number

$\text{Var}(X) = \frac{1}{12}(21^2 - 1) = \frac{440}{12} = \frac{110}{3}$ $(= 36.7)$

9 (c) Values of $Y \leqslant 7$: 5.6, 6.0, 6.6, 6.9, so 4 out of 25

$P(Y \leqslant 7) \approx \frac{4}{25} = 0.16$

9 (d) $\text{Var}(Y) = \frac{\text{Var}(X)}{10}$, so $\text{Var}(Y)$ is much smaller than $\text{Var}(X)$.

$E(Y) = E(X) = 10$, so values of Y are more likely to be close to the mean 10 than values of X, so $P(Y \leqslant 7) < P(X \leqslant 7)$.

9 (e) $E(W) = 10$, $\text{Var}(W) = \frac{110/3}{30} = \frac{11}{9}$

By the CLT, $W \approx N(10, \frac{11}{9})$

The total of the 30 values is an integer, so $W \leqslant 7$ means total $\leqslant 210$; continuity correction: total < 210.5, $W < 7\frac{1}{60}$

$P(W \leqslant 7) \approx P\left(Z < \frac{7.01667 - 10}{\sqrt{11/9}}\right) = P(Z < -2.6985)$

$= 0.00348$ (3 s.f.)

10 (a) $\text{Expected} = \frac{\text{row total} \times \text{column total}}{\text{grand total}}$

Low, normal $= \frac{32 \times 12}{60} = 6.4$, Low, high $= \frac{28 \times 12}{60} = 5.6$

Question number

Medium, normal $= \frac{32 \times 31}{60} = 16.5333$

Medium, high $= \frac{28 \times 31}{60} = 14.4667$

	Low	Medium	High
Normal	6.4	16.5333	9.0667
High	5.6	14.4667	7.9333

10 (b) Contribution $= \frac{(O - E)^2}{E} = \frac{(5 - 9.0667)^2}{9.0667}$

$= 1.8240$ (4 d.p.)

10 (c) H_0: there is no association between dietary fat intake and cholesterol level

H_1: there is some association between dietary fat intake and cholesterol level

$X^2 = 1.0563 + 0.1301 + 1.8240 + 1.2071 + 0.1487 + 2.0846$

$X^2 = 6.4508$

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

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

$6.4508 > 5.991$, so the result is significant: reject H_0.

There is sufficient evidence at the 5% level to suggest that there is some association between dietary fat intake and cholesterol level.

10 (d) Low fat: contributions 1.0563 and 1.2071 show more people than expected have normal cholesterol and fewer than expected have high cholesterol.

Medium fat: small contributions (0.1301, 0.1487), so the numbers

Question number

with normal and high cholesterol are about as expected.

High fat: contributions 1.8240 and 2.0846 show fewer people than expected have normal cholesterol and more than expected have high cholesterol.

11 (a) A t test requires the population to be Normally distributed.

The Normal probability plot is not close to a straight line and the K-S p-value 0.049 < 0.05 is low; both suggest the data do not come from a Normal distribution.

So a Wilcoxon signed rank test should be used, since it does not require the population to be Normally distributed.

11 (b) H_0: the population median increase in lean body mass is 1 kg

H_1: the population median increase in lean body mass is less than 1 kg

Value	Value − 1	|Value − 1|	Rank
−0.84	−1.84	1.84	10
−0.76	−1.76	1.76	9
−0.16	−1.16	1.16	8
0.43	−0.57	0.57	4
1.31	0.31	0.31	1
1.32	0.32	0.32	2
1.47	0.47	0.47	3
1.64	0.64	0.64	5
1.93	0.93	0.93	6
2.14	1.14	1.14	7

$W_- = 10 + 9 + 8 + 4 = 31$

$W_+ = 1 + 2 + 3 + 5 + 6 + 7 = 24$

Test statistic = $W_+ = 24$

Question number

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

$24 > 10$, so the result is not significant: do not reject H_0.

There is insufficient evidence at the 5% level to suggest that the supplement increases lean body mass by less than 1 kg on average, so insufficient evidence that the researcher's belief is correct.

11 (c)

A (single sample) t test could have been used.

H_0: $\mu = 1$, H_1: $\mu < 1$

where μ kg is the population mean increase in lean body mass.

12 (a)

$F(a) = 1$: $k(a \times a - 0.5a^2) = 1$

$0.5ka^2 = 1 \Rightarrow k = \frac{2}{a^2}$

Median m: $F(m) = 0.5$

$\frac{2}{a^2}(am - 0.5m^2) = 0.5$

$4am - 2m^2 = a^2$

$2m^2 - 4am + a^2 = 0$

$m = \frac{4a \pm \sqrt{16a^2 - 8a^2}}{4} = \frac{4a \pm 2\sqrt{2}a}{4} = a(1 \pm \frac{\sqrt{2}}{2})$

$m = a(1 + \frac{\sqrt{2}}{2}) > a$ so is rejected (m must be in $[0, a]$)

Median $= a(1 - \frac{\sqrt{2}}{2})$ $(\approx 0.293a)$

12 (b)

$a = 10$: $k = \frac{2}{100} = \frac{1}{50}$, $F(x) = \frac{1}{50}(10x - 0.5x^2)$

Question number

$f(x) = F'(x) = \frac{1}{50}(10 - x)$ for $0 \leq x \leq 10$

$E(X) = \int_0^{10} \frac{1}{50}x(10 - x)\,dx = \frac{1}{50}\left[5x^2 - \frac{x^3}{3}\right]_0^{10}$

$= \frac{1}{50}\left(500 - \frac{1000}{3}\right) = \frac{10}{3}$

$E(X^2) = \int_0^{10} \frac{1}{50}x^2(10 - x)\,dx = \frac{1}{50}\left[\frac{10x^3}{3} - \frac{x^4}{4}\right]_0^{10}$

$= \frac{1}{50}\left(\frac{10000}{3} - 2500\right) = \frac{50}{3}$

$Var(X) = \frac{50}{3} - \left(\frac{10}{3}\right)^2 = \frac{50}{9}, \quad \sigma = \frac{5\sqrt{2}}{3} = 2.3570$

$\mu - \sigma = 0.97631, \quad \mu + \sigma = 5.69036$

$P(0.97631 < X < 5.69036) = F(5.69036) - F(0.97631)$

$= \frac{1}{50}(56.9036 - 0.5 \times 5.69036^2) - \frac{1}{50}(9.7631 - 0.5 \times 0.97631^2)$

$= 0.81427 - 0.18573$

$= 0.6285$ (4 d.p.) (exactly $\frac{4\sqrt{2}}{9}$)