OCR
Oxford Cambridge and RSA

Thursday 7 October 2021 – Afternoon

A Level Further Mathematics B (MEI)

Y422/01 Statistics Major

Time allowed: 2 hours 15 minutes

9 0 5 0 6 1 9 3 7 6

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 2021 [H/508/5594]
DC (ST/FC) 300437/4
OCR is an exempt Charity

stemup.co.uk

Section A (29 marks)

Answer **all** the questions.

1 When babies are born, their head circumferences are measured. A random sample of 50 newborn female babies is selected. The sample mean head circumference is 34.711 cm. The sample standard deviation head circumference is 1.530 cm.

(a) Determine a 95% confidence interval for the population mean head circumference of newborn female babies. **[4]**

(b) Explain why you can calculate this interval even though the distribution of the population of head circumferences of newborn female babies is unknown. **[2]**

2 In a game at a charity fair, a player rolls 3 unbiased six-sided dice. The random variable X represents the difference between the highest and lowest scores.

(a) Show that $P(X = 0) = \frac{1}{36}$. **[2]**

The table shows the probability distribution of X.

r	0	1	2	3	4	5
$P(X = r)$	$\frac{1}{36}$	$\frac{5}{36}$	$\frac{2}{9}$	$\frac{1}{4}$	$\frac{2}{9}$	$\frac{5}{36}$

(b) Draw a graph to illustrate the distribution. **[2]**

(c) Describe the shape of the distribution. **[1]**

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

Find each of the following.

- $E(X)$
- $Var(X)$ **[5]**

As a result of playing the game, the player receives $30X$ pence from the organiser of the game.

(e) Find the variance of the amount that the player receives. **[1]**

(f) The player pays k pence to play the game.

Given that the average profit made by the organiser is 12.5 pence per game, determine the value of k. **[2]**

© OCR 2021

stemup.co.uk

3 In air traffic management, air traffic controllers send radio messages to pilots. On receiving a message, the pilot repeats it back to the controller to check that it has been understood correctly.

At a particular site, on average 4% of messages sent by controllers are not repeated back correctly and so have been misunderstood. You should assume that instances of messages being misunderstood occur randomly and independently.

(a) Find the probability that exactly 2 messages are misunderstood in a sequence of 50 messages. **[2]**

(b) Find the probability that in a sequence of messages, the 10th message is the first one which is misunderstood. **[1]**

(c) Find the probability that in a sequence of 20 messages, there are no misunderstood messages. **[1]**

(d) Determine the expected number of messages required for 3 of them to be misunderstood. **[3]**

(e) Determine the probability that in a sequence of messages, the 3rd misunderstood message is the 60th message in the sequence. **[3]**

© OCR 2021

stemup.co.uk

Section B (91 marks)

Answer **all** the questions.

4 A radioactive source contains 1 000 000 nuclei of a particular radioisotope. On average 1 in 200 000 of these nuclei will decay in a period of 1 second. The random variable X represents the number of nuclei which decay in a period of 1 second. You should assume that nuclei decay randomly and independently of each other.

(a) Explain why you could use either a binomial distribution or a Poisson distribution to model the distribution of X. **[3]**

Use a Poisson distribution to answer parts **(b)** and **(c)**.

(b) Calculate each of the following probabilities.

- $P(X = 6)$
- $P(X > 6)$ **[3]**

(c) Determine an estimate of the probability that at least 60 nuclei decay in a period of 10 seconds. **[2]**

© OCR 2021

stemup.co.uk

5 A manufacturer uses three types of capacitor in a particular electronic device. The capacitances, measured in suitable units, are modelled by independent Normal distributions with means and standard deviations as shown in the table.

	Capacitance	
Type	Mean	Standard deviation
A	3.9	0.32
B	7.8	0.41
C	30.2	0.64

(a) Determine the probability that the total capacitance of a randomly chosen capacitor of Type B and two randomly chosen capacitors of Type A is at least 16 units. **[3]**

(b) Determine the probability that the capacitance of a randomly chosen capacitor of Type C is within 1 unit of the total capacitance of four randomly chosen capacitors of Type B. **[3]**

When the manufacturer gets a new batch of 1000 capacitors from the supplier, a random sample of 10 of them is tested to check the capacitances. For a new batch of Type C capacitors, summary statistics for the capacitances, x units, of the random sample are as follows.

$n = 10 \qquad \sum x = 299.6 \qquad \sum x^2 = 8981.0$

You should assume that the capacitances of the sample come from a Normally distributed population, but you should **not** assume that the standard deviation is 0.64 as for previous Type C capacitors.

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

Carry out a hypothesis test at the 5% significance level to check whether it is reasonable to assume that the capacitors in this batch have the specified mean capacitance for Type C of 30.2 units. **[11]**

stemup.co.uk

6 Cosmic rays passing through the upper atmosphere cause muons, and other types of particle, to be formed. Muons can be detected when they reach the surface of the earth. It is known that the mean number of muons reaching a particular detector is 1.7 per second. The numbers of muons reaching this detector in 200 randomly selected periods of 1 second are shown in **Fig. 6.1**.

Number of muons	0	1	2	3	4	5	6	⩾7
Frequency	34	65	55	24	14	6	2	0

Fig. 6.1

(a) Use the values of the sample mean and sample variance to discuss the suitability of a Poisson distribution as a model. **[3]**

The screenshot in **Fig. 6.2** shows part of a spreadsheet to assess the goodness of fit of the distribution Po(1.7).

	A	B	C	D	E
1	Number of muons	Observed frequency	Poisson probability	Expected frequency	Chi-squared contribution
2	0	34	0.1827	36.5367	0.1761
3	1	65			
4	2	55	0.2640	52.7955	0.0920
5	3	24	0.1496	29.9175	1.1704
6	4	14			0.1299
7	⩾5	8	0.0296	5.9230	0.7284

Fig. 6.2

(b) Calculate the missing values in each of the following cells.

- C3
- D3
- E3

[4]

(c) Explain why the numbers for 5, 6 and at least 7 muons have been combined into the single category of at least 5 muons, as shown in **Fig. 6.2**. **[1]**

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

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

© OCR 2021

stemup.co.uk

7 A physiotherapist is investigating hand grip strength in adult women under 30 years old. She thinks that the grip strength of the dominant hand will be on average 2 kg higher than the grip strength of the non-dominant hand.

The physiotherapist selects a random sample of 12 adult women under 30 years old and measures the grip strength of each of their hands. She then uses software to produce a 95% confidence interval for the mean difference in grip strength between the two hands (dominant minus non-dominant), as shown in **Fig. 7**.

T Estimate of a Mean

Confidence Level 0.95

Sample

Mean 2.79

s 3.92

N 12

Result

T Estimate of a Mean

Mean	2.79
s	3.92
SE	1.13161
N	12
df	11
Lower Limit	0.29935
Upper Limit	5.28065
Interval	2.79 ± 2.49065

Fig. 7

(a) Explain why the physiotherapist used the same people for testing their dominant and non-dominant grip strengths. **[2]**

(b) State any assumptions necessary in order to construct the confidence interval shown in **Fig. 7**. **[2]**

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

(d) The physiotherapist then finds some data which have previously been collected on grip strength using a sample of 100 adult women. A 95% confidence interval, based on this sample and calculated using a Normal distribution, for the mean difference in grip strength between the two hands (dominant minus non-dominant) is (1.94, 2.84).

(i) For this sample, find

- the mean difference
- the standard deviation of the differences. **[3]**

(ii) Explain what you would need to know about the nature of this sample if you wanted to draw conclusions about the mean difference in grip strength in the population of adult women. **[2]**

© OCR 2021

stemup.co.uk

8 **(a)** VO_{2max} is a measure of athletic fitness. Since VO_{2max} is fairly time-consuming and expensive to measure, an exercise scientist wants to predict VO_{2max} from data such as times for running different distances. The scientist uses these data for a random sample of 15 athletes to predict their VO_{2max} values, denoted by y, in suitable units. She also obtains accurate measurements of the VO_{2max} values, denoted by x, in the same units.

The scatter diagram in **Fig. 8.1** shows the values of x and y obtained, together with the equation of the regression line of y on x and the value of r^2.

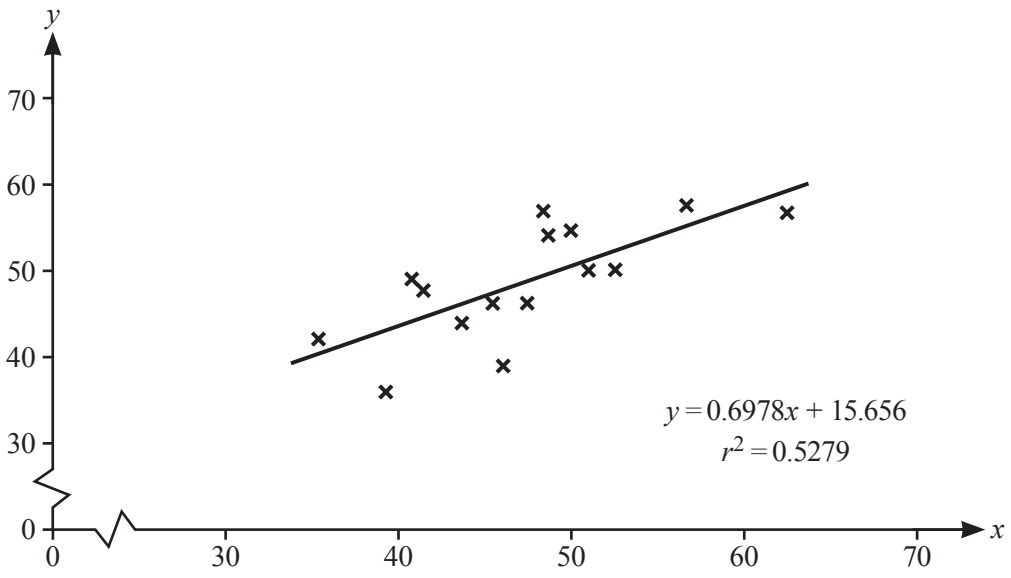


Fig. 8.1

(i) Use the regression line to estimate the predicted VO_{2max} of an athlete whose accurately measured VO_{2max} is 50. **[1]**

(ii) Comment on the reliability of your estimate. **[2]**

(iii) The equation of the regression line of x on y is $x = 0.7565y + 10.493$.

Find the coordinates of the point at which the two regression lines meet. **[1]**

(iv) State what the point you found in part **(iii)** represents. **[1]**

© OCR 2021

stemup.co.uk

(b) It is known that there is negative correlation between VO_{2max} and marathon times in very good runners (those whose best marathon times are under 3 hours). The exercise scientist wishes to know whether the same applies to runners who take longer to run a marathon. She selects a random sample of 20 runners whose best marathon times are between $3\frac{1}{2}$ hours and $4\frac{1}{2}$ hours and accurately measures their VO_{2max}.

Fig. 8.2 is a scatter diagram of accurately measured VO_{2max}, v units, against best marathon time, t hours, for these runners.

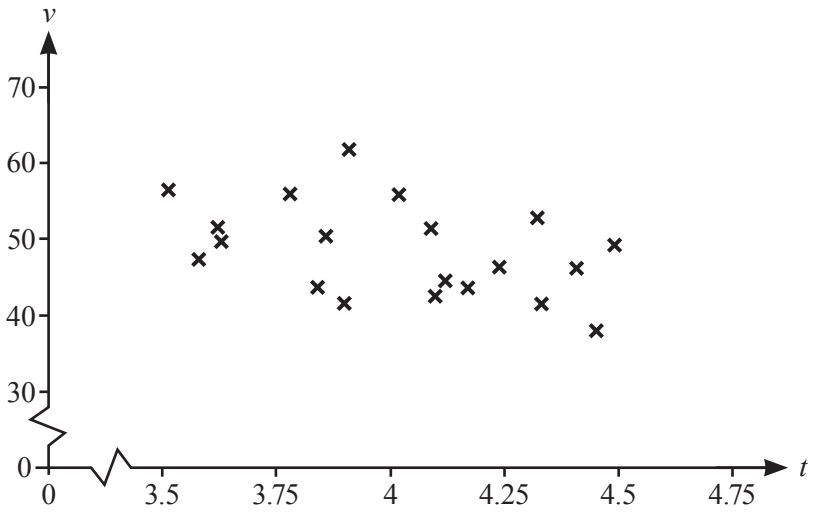


Fig. 8.2

(i) Explain why the exercise scientist comes to the conclusion that a test based on Pearson's product moment correlation coefficient may be valid. **[2]**

Summary statistics for the 20 runners are as follows.

$\sum t = 80.37$ $\sum v = 970.86$ $\sum t^2 = 324.71$ $\sum v^2 = 47\,829.24$ $\sum tv = 3886.53$

(ii) Find the value of Pearson's product moment correlation coefficient. **[4]**

(iii) Carry out a test at the 5% significance level to investigate whether there is negative correlation between accurately measured VO_{2max} and best marathon time for runners whose best marathon times are between $3\frac{1}{2}$ hours and $4\frac{1}{2}$ hours. **[5]**

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

(a) Given that n is odd, determine $P(X > \frac{1}{2}n)$, giving your answer as a single fraction in terms of n. **[3]**

(b) Determine the variance of the sum of 10 independent values of X, giving your answer in the form $an^2 + bn$, where a and b are constants. **[3]**

stemup.co.uk

10 Sarah takes a bus to work each weekday morning and returns each evening. The times in minutes that she has to wait for the bus in the morning and evening are modelled by uniform distributions over the intervals [0, 10] and [0, 6] respectively. The times in minutes for the bus journeys in the morning and evening are modelled by N(25, 4) and N(28, 16) respectively. You should assume that all of the times are independent. The total time in minutes that she takes for her two journeys, including the waiting times, is denoted by the random variable T.

The spreadsheet below shows the first 20 rows of a simulation of 500 return journeys. It also shows in column H the numbers of values of T that are less than or equal to the corresponding values in column G. For example, there are 156 out of the 500 simulated values of T which are less than or equal to 58 minutes. All of the times have been rounded to 2 decimal places.

	A	B	C	D	E	F	G	H
1	**Waiting time morning**	**Journey time morning**	**Waiting time evening**	**Journey time evening**	**Total time**		**Total time t**	**Number $\leqslant t$**
2	0.89	20.78	1.88	26.30	49.86		46	0
3	3.55	21.24	1.04	29.61	55.44		48	4
4	2.13	21.83	2.40	28.64	55.00		50	13
5	5.12	25.04	3.13	24.30	57.60		52	35
6	4.03	27.49	2.19	30.81	64.52		54	57
7	2.47	20.54	4.32	34.61	61.93		56	104
8	3.21	26.93	3.78	27.66	61.58		58	156
9	9.72	24.15	0.63	27.53	62.03		60	218
10	1.59	28.45	0.08	35.87	65.99		62	288
11	7.34	23.04	4.02	24.77	59.17		64	357
12	1.04	24.69	1.66	31.95	59.33		66	408
13	7.17	22.16	2.55	25.39	57.28		68	441
14	5.20	26.97	2.41	30.05	64.62		70	475
15	5.01	26.84	1.88	36.21	69.93		72	490
16	3.76	26.03	2.21	30.96	62.96		74	496
17	0.96	23.72	2.55	29.36	56.59		76	500
18	8.64	24.97	2.82	26.39	62.82			
19	0.59	20.82	4.57	31.41	57.38			
20	9.85	23.68	5.54	29.92	68.99			

(a) Use the spreadsheet output to estimate each of the following.

- $P(T \leqslant 56)$
- $P(T > 61)$ **[2]**

(b) The random variable W is Normally distributed with the same mean and variance as T. Find each of the following.

- $P(W \leqslant 56)$
- $P(W > 61)$ **[5]**

(c) Explain why, if many more journeys were used in the simulation, you would expect $P(T > 61)$ to be extremely close to $P(W > 61)$. **[2]**

© OCR 2021

stemup.co.uk

11 The continuous random variable X has probability density function given by

$$f(x) = \begin{cases} ax^2 & 0 \leqslant x < 2, \\ b(3-x)^2 & 2 \leqslant x \leqslant 3, \\ 0 & \text{otherwise,} \end{cases}$$

where a and b are positive constants.

(a) Given that $E(X) = 2$, determine the values of a and b. **[5]**

(b) Determine the median value of X. **[3]**

(c) A random sample of 50 observations of X is selected.

Given that $\text{Var}(X) = 0.2$, determine an estimate of the probability that the mean value of the 50 observations is less than 1.9. **[3]**

END OF QUESTION PAPER

© OCR 2021

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 the Cambridge Assessment Group; Cambridge Assessment is the brand name of University of Cambridge Local Examinations Syndicate (UCLES), which is itself a department of the University of Cambridge.

© OCR 2021

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)

$\bar{x} = 34.711, \quad s = 1.530, \quad n = 50$

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

$= 34.711 \pm 1.96 \times \frac{1.530}{\sqrt{50}}$

$= 34.711 \pm 0.424$

(34.287, 35.135) cm

1 (b)

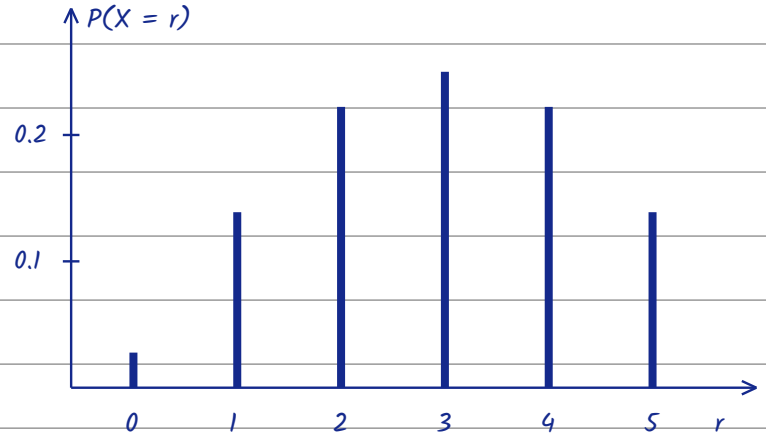
$n = 50$ is a large sample, so by the Central Limit Theorem the distribution of the sample mean is approximately Normal, whatever the distribution of the population.

2 (a)

$X = 0$ means all three dice show the same score.

$P(X = 0) = \frac{6}{6} \times \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$ as required

2 (b)



Question number

2 (c) The distribution has slight negative skew (it is unimodal).

2 (d) $E(X) = 0 \times \frac{1}{36} + 1 \times \frac{5}{36} + 2 \times \frac{2}{9} + 3 \times \frac{1}{4} + 4 \times \frac{2}{9} + 5 \times \frac{5}{36}$

$= \frac{0 + 5 + 16 + 27 + 32 + 25}{36} = \frac{105}{36} = \frac{35}{12}$ $(= 2.917)$

$E(X^2) = 0 \times \frac{1}{36} + 1 \times \frac{5}{36} + 4 \times \frac{2}{9} + 9 \times \frac{1}{4} + 16 \times \frac{2}{9} + 25 \times \frac{5}{36}$

$= \frac{0 + 5 + 32 + 81 + 128 + 125}{36} = \frac{371}{36}$ $(= 10.306)$

$Var(X) = E(X^2) - [E(X)]^2 = \frac{371}{36} - (\frac{35}{12})^2$

$= \frac{1484}{144} - \frac{1225}{144} = \frac{259}{144}$ $(= 1.80)$

2 (e) $Var(30X) = 30^2 \times Var(X) = 900 \times \frac{259}{144} = 1618.75$

$= 1620$ pence^2 (3 s.f.)

2 (f) Average amount received $= E(30X) = 30 \times \frac{35}{12} = 87.5$ pence

Average profit: $k - 87.5 = 12.5$

$k = 100$

3 (a) Y = number misunderstood in 50 messages, $Y \sim B(50, 0.04)$

$P(Y = 2) = {}^{50}C_2 \times 0.04^2 \times 0.96^{48} = 0.276$ (3 s.f.)

3 (b) $P = 0.96^9 \times 0.04 = 0.0277$ (3 s.f.)

Question number

3 (c) $P = 0.96^{20} = 0.442$ (3 s.f.)

3 (d) Number of messages up to and including the first misunderstood one is geometric, Geo(0.04), with expected value $\frac{1}{0.04} = 25$

The numbers of messages until each of the 3 misunderstood ones are independent geometric variables, so expected number = 25 + 25 + 25 = 75

3 (e) Need exactly 2 misunderstood in the first 59 messages, then the 60th message misunderstood.

Using B(59, 0.04): P(2 in first 59) = 0.2672

$P = 0.2672 \times 0.04 = 0.0107$ (3 s.f.)

4 (a) Each nucleus decays randomly and independently with constant probability $\frac{1}{200000}$, and X is the number out of a fixed number 1 000 000 that decay, so $X \sim B(1000000, \frac{1}{200000})$.

Since n = 1 000 000 is large and $p = \frac{1}{200000}$ is small, the Poisson distribution Po(np) = Po(5) is also appropriate.

4 (b) $\lambda = 1000000 \times \frac{1}{200000} = 5$, $X \sim Po(5)$

$P(X = 6) = \frac{e^{-5} \times 5^6}{6!} = 0.146$ (3 s.f.)

Question number

$P(X > 6) = 1 - P(X \leq 6) = 1 - 0.7622 = 0.238$ (3 s.f.)

4 (c) Mean in 10 seconds $= 10 \times 5 = 50$, $Y \sim Po(50)$

$P(Y \geq 60) = 1 - P(Y \leq 59) = 1 - 0.9077 = 0.0923$ (3 s.f.)

5 (a) $S = A_1 + A_2 + B$

$E(S) = 2 \times 3.9 + 7.8 = 15.6$

$Var(S) = 2 \times 0.32^2 + 0.41^2 = 0.2048 + 0.1681 = 0.3729$

$S \sim N(15.6, 0.3729)$

$P(S \geq 16) = P\left(Z \geq \frac{16 - 15.6}{\sqrt{0.3729}}\right) = P(Z \geq 0.6550)$

$= 0.256$ (3 s.f.)

5 (b) $D = B_1 + B_2 + B_3 + B_4 - C$

$E(D) = 4 \times 7.8 - 30.2 = 1$

$Var(D) = 4 \times 0.41^2 + 0.64^2 = 0.6724 + 0.4096 = 1.082$

$D \sim N(1, 1.082)$

$P(-1 < D < 1) = P\left(\frac{-1 - 1}{\sqrt{1.082}} < Z < 0\right) = P(-1.9227 < Z < 0)$

$= 0.473$ (3 s.f.)

5 (c) $H_0: \mu = 30.2$, $H_1: \mu \neq 30.2$

where μ is the population mean capacitance of the batch

$\bar{x} = \frac{299.6}{10} = 29.96$

$s^2 = \frac{1}{9}\left(8981.0 - \frac{299.6^2}{10}\right) = \frac{4.984}{9} = 0.5538$, $s = 0.7442$

Population variance unknown and population Normal, so use t_9

Question number

Test statistic $t = \frac{29.96 - 30.2}{\sqrt{0.5538 / 10}} = \frac{-0.24}{0.2353} = -1.020$

Two-tailed 5% critical value for $t_9 = 2.262$

$-1.020 > -2.262$, so the result is not significant: do not reject H_0.

There is insufficient evidence at the 5% level to suggest that the mean capacitance of this batch is different from 30.2 units, so it is reasonable to assume the specified mean of 30.2 units.

6 (a) $\Sigma fx = 345, \quad \Sigma fx^2 = 947, \quad n = 200$

$\text{Mean} = \frac{345}{200} = 1.725$

$\text{Variance} = \frac{1}{199}\left(947 - \frac{345^2}{200}\right) = \frac{351.875}{199} = 1.768$

For a Poisson distribution mean = variance. The sample variance is reasonably close to the sample mean, so this supports the suitability of a Poisson model.

6 (b) $C3 = P(X = 1) = e^{-1.7} \times 1.7 = 0.3106$

$D3 = 200 \times 0.31056 = 62.1124$

$E3 = \frac{(65 - 62.1124)^2}{62.1124} = 0.1342$

6 (c) Otherwise some expected frequencies would be less than 5, which is too small for the chi-squared test to be valid.

6 (d) H_0: Po(1.7) is a good model for the number of muons per second

H_1: Po(1.7) is not a good model for the number of muons per second

$X^2 = 0.1761 + 0.1342 + 0.0920 + 1.1704 + 0.1299 + 0.7284$

Question number	
	$X^2 = 2.431$
	6 cells, no parameters estimated: degrees of freedom = 6 − 1 = 5
	5% critical value of $\chi^2_5 = 11.07$
	2.431 < 11.07, so the result is not significant: do not reject H_0.
	There is insufficient evidence at the 5% level to suggest that
	the Po(1.7) model is not a good fit.
7 (a)	Pairing removes the differences in grip strength between
	different people, so the only differences compared are between
	the dominant and non-dominant hands of the same person.
7 (b)	The population of differences in grip strength (dominant minus
	non-dominant) must be Normally distributed.
7 (c)	Yes, it supports her belief because the interval (0.299, 5.281)
	contains 2.
7 (d)(i)	Mean difference = $\frac{1.94 + 2.84}{2}$ = 2.39 kg
	Half-width = $\frac{2.84 - 1.94}{2}$ = 0.45
	$1.96 \times \frac{s}{\sqrt{100}} = 0.45$
	$s = \frac{0.45 \times 10}{1.96}$ = 2.30 kg (3 s.f.)
7 (d)(ii)	The sample must be a random sample of adult women, since only
	a random sample allows valid inferences to be made about the

Question number	
	population.
8 (a)(i)	$y = 0.6978 \times 50 + 15.656 = 50.5$
8 (a)(ii)	$x = 50$ lies within the range of the data (interpolation), but the points are not very close to the line and $r^2 = 0.5279$ is not close to 1, so the estimate is only moderately reliable.
8 (a)(iii)	$y = 0.6978x + 15.656$ and $x = 0.7565y + 10.493$ $x = 0.7565(0.6978x + 15.656) + 10.493$ $x - 0.52789x = 11.8438 + 10.493 \Rightarrow 0.47211x = 22.3368$ $x = 47.31$, $y = 0.6978 \times 47.31 + 15.656 = 48.67$ Coordinates (47.3, 48.7)
8 (a)(iv)	It is the point $(\bar{x}, \bar{y})$, whose coordinates are the means of the x-values and the y-values (the centroid).
8 (b)(i)	The scatter diagram is very roughly elliptical, so the data may come from a bivariate Normal distribution.
8 (b)(ii)	$S_{tv} = 3886.53 - \frac{80.37 \times 970.86}{20} = -14.8709$ $S_{tt} = 324.71 - \frac{80.37^2}{20} = 1.74316$ $S_{vv} = 47829.24 - \frac{970.86^2}{20} = 700.783$ $r = \frac{S_{tv}}{\sqrt{S_{tt}S_{vv}}} = \frac{-14.8709}{\sqrt{1.74316 \times 700.783}}$

Question number

$r = -0.4255$ (4 s.f.)

8 (b)(iii) H_0: $\rho = 0$, H_1: $\rho < 0$

where ρ is the population pmcc between best marathon time and VO_2max for runners with best times between 3½ and 4½ hours

For $n = 20$, 5% one-tailed critical value = 0.3783

$-0.4255 < -0.3783$, so the result is significant: reject H_0.

There is sufficient evidence at the 5% level to suggest that there is negative correlation between VO_2max and best marathon time for these runners.

9 (a) X takes the $2n + 1$ values $-n, \ldots, 0, \ldots, n$, each with probability $\frac{1}{2n+1}$

n odd, so $\frac{1}{2}n$ is not an integer and $X > \frac{1}{2}n$ means

$X = \frac{n+1}{2}, \ldots, n$: that is $n - \frac{n+1}{2} + 1 = \frac{n+1}{2}$ values

$P(X > \frac{1}{2}n) = \frac{(n+1)/2}{2n+1} = \frac{n+1}{2(2n+1)}$

9 (b) X is uniform on $2n + 1$ consecutive integers, so

$\text{Var}(X) = \frac{1}{12}[(2n+1)^2 - 1] = \frac{4n^2 + 4n}{12} = \frac{n^2 + n}{3}$

$\text{Var}(X_1 + \ldots + X_{10}) = 10\,\text{Var}(X)$ (independent)

$= \frac{10}{3}n^2 + \frac{10}{3}n$

Question number

10 (a) $P(T \leqslant 56) \approx \frac{104}{500} = 0.208$

$P(T \leqslant 61) \approx \frac{1}{2}(218 + 288) \div 500 = \frac{253}{500}$

$P(T > 61) \approx 1 - \frac{253}{500} = 0.494$

10 (b) $E(T) = 5 + 25 + 3 + 28 = 61$

$Var(T) = \frac{1}{12} \times 10^2 + 4 + \frac{1}{12} \times 6^2 + 16$

$= \frac{100}{12} + \frac{36}{12} + 20 = \frac{94}{3} \quad (= 31.333)$

$W \sim N(61, \frac{94}{3})$

$P(W \leqslant 56) = P(Z \leqslant \frac{56 - 61}{\sqrt{31.333}}) = P(Z \leqslant -0.8932) = 0.186$ (3 s.f.)

$P(W > 61) = 0.5$ (61 is the mean of W)

10 (c) The mean of T is 61 and the uniform and Normal distributions are all symmetrical, so T is symmetrical about 61. So with many more journeys the simulated $P(T > 61)$ would be very close to 0.5, which is $P(W > 61)$.

11 (a) $\int_0^2 ax^2 \, dx + \int_2^3 b(3 - x)^2 \, dx = 1$

$\left[\frac{ax^3}{3}\right]_0^2 + \left[-\frac{b(3 - x)^3}{3}\right]_2^3 = 1 \quad \Rightarrow \quad \frac{8}{3}a + \frac{1}{3}b = 1$

Question number

$E(X) = \int_0^2 ax^3\,dx + \int_2^3 bx(3-x)^2\,dx = 2$

$\int_2^3 x(3-x)^2\,dx = \int_2^3 (9x - 6x^2 + x^3)\,dx = \left[\frac{9x^2}{2} - 2x^3 + \frac{x^4}{4}\right]_2^3 = \frac{3}{4}$

$\left[\frac{ax^4}{4}\right]_0^2 + \frac{3}{4}b = 2 \quad \Rightarrow \quad 4a + \frac{3}{4}b = 2$

From the first equation $b = 3 - 8a$:

$4a + \frac{3}{4}(3 - 8a) = 2 \Rightarrow 16a + 9 - 24a = 8 \Rightarrow a = \frac{1}{8}$

$a = \frac{1}{8}, \quad b = 3 - 1 = 2$

11 (b)

$F(2) = \int_0^2 \frac{1}{8}x^2\,dx = \frac{8}{24} = \frac{1}{3} < \frac{1}{2}$, so the median m is in $[2, 3]$

$\frac{1}{3} + \int_2^m 2(3-x)^2\,dx = \frac{1}{2}$

$\left[-\frac{2}{3}(3-x)^3\right]_2^m = \frac{1}{6} \quad \Rightarrow \quad -\frac{2}{3}(3-m)^3 + \frac{2}{3} = \frac{1}{6}$

$(3-m)^3 = \frac{3}{4}$

$m = 3 - \sqrt[3]{0.75} = 2.09$ (3 s.f.)

11 (c)

$n = 50$ is large, so by the Central Limit Theorem

$\bar{X} \approx N\left(2, \frac{0.2}{50}\right) = N(2, 0.004)$

$P(\bar{X} < 1.9) = P\left(Z < \frac{1.9 - 2}{\sqrt{0.004}}\right) = P(Z < -1.5811)$

$= 0.0569$ (3 s.f.)