UNIVERSITY OF CAMBRIDGE INTERNATIONAL EXAMINATIONS
Cambridge International Level 3 Pre-U Certificate
Short Course

MATHEMATICS (STATISTICS WITH PURE MATHEMATICS) **1347/01**

Paper 1 Pure Mathematics **For Examination from 2012**

SPECIMEN PAPER

1 hour 45 minutes

Additional Materials: Answer Booklet/Paper
Graph Paper
List of Formulae (MF21)

READ THESE INSTRUCTIONS FIRST

If you have been given an Answer Booklet, follow the instructions on the front cover of the Booklet.
Write your Centre number, candidate number and name on all the work you hand in.
Write in dark blue or black pen.
You may use a soft pencil for any diagrams or graphs.
Do not use staples, paper clips, highlighters, glue or correction fluid.

Answer **all** the questions.
Give non-exact numerical answers correct to 3 significant figures, or 1 decimal place in the case of angles in degrees, unless a different level of accuracy is specified in the question.
The use of an electronic calculator is expected, where appropriate.
You are reminded of the need for clear presentation in your answers.

At the end of the examination, fasten all your work securely together.
The number of marks is given in brackets [ ] at the end of each question or part question.
The total number of marks for this paper is 65.

This document consists of **3** printed pages and **1** blank page.

UNIVERSITY of CAMBRIDGE International Examinations

[Turn over

1 A farmer has a rectangular field. The length of the field is 10 metres more than its width and the area of the field is 39 square metres. The farmer writes this as $x(x + 10) = 39$.

(i) By completing the square, express $x(x + 10) = 39$ in the form $(x + a)^2 = b$, where a and b are constants to be found. [2]

(ii) Hence write down the width and length of the field. [1]

2 The sensitivity, S (dB), of a loudspeaker at different acoustic efficiency percentages, E (%), is given in the table.

E	1	2	5	10	20
S	92	95	99	102	105

It is claimed that there is a relationship between E and S of the form $S = kE^n$.

When the values of $\ln S$ are plotted against the values of $\ln E$ the points approximately fit a straight line with gradient 0.0445.

(i) Deduce values for k and n. [3]

(ii) What sensitivity would you expect from a loudspeaker with 15% efficiency? [2]

3 Jack needs to know the value of $(2.03)^5$ but he cannot be bothered to find his calculator.

(i) Write down and simplify the expansion of $(2 + 3x)^5$ in ascending powers of x as far as the term in x^2. [4]

(ii) Use your answer to part **(i)** to calculate an approximate value for $(2.03)^5$ and show that it agrees with the value from your calculator to 2 decimal places. [2]

4 **(i)** Show that at the points of intersection of $y = 3x^2 - 2x + 2$ and $y = 4x + k$ the x-values must satisfy the quadratic equation $3x^2 - 6x + (2 - k) = 0$. [1]

(ii) Calculate the discriminant of $3x^2 - 6x + (2 - k)$ and hence find the point on the curve $y = 3x^2 - 2x + 2$ where the tangent is parallel to the line joining $(-1, 7)$ and $(3, 23)$. [5]

5 A triangle is formed by joining the points $A\,(1, 2)$, $B\,(5, 5)$ and $C\,(2, 9)$.

(i) Calculate the length of the line segment AC, giving your answer in surd form. [1]

(ii) Calculate the gradient of AB and the gradient of BC. What can you deduce about the line segments AB and BC? [3]

(iii) Give the centre of the circle that passes through A, B and C. Also give the radius of the circle, expressing your answer in terms of $\sqrt{2}$. [3]

6 In his Anthropology coursework, Tom is trying to construct a function to model the cross-sectional area, y, of a small ceramic pot in terms of its depth, x. He needs the graph of the function to have a minimum turning point at (0, 8). He has decided to try a function of the form $y = -2x^4 + ax^2 + bx + 8$.

(i) Show that for the function to have a turning point at (0, 8), b must equal 0. Calculate $\frac{d^2y}{dx^2}$ and show that setting $a > 0$ will guarantee that the turning point is a minimum turning point. [4]

(ii) Find, in terms of a, the x-values of the other stationary points on the graph of Tom's function. [3]

Tom decides to use the function $y = -2x^4 + 6x^2 + 8$. He then realises that he has recorded the x-values in his coursework in centimetres instead of metres and the y-values in cm^2 instead of m^2.

(iii) Explain why Tom should change his function to $y = -20\,000x^4 + 6x^2 + 0.0008$. [2]

The function $y = -20\,000x^4 + 6x^2 + 0.0008$ gives the cross-sectional area, $y\,m^2$, as a function of the depth, $x\,m$, where $0 \leqslant x \leqslant 0.02$.

(iv) Use integration to calculate the volume of the ceramic pot in m^3. [4]

7 **(i)** The graph of the quadratic polynomial $y = (x - A)^2 + B$ has its minimum turning point at (−1, 7). Show that the point (3, 23) also lies on the curve. [3]

(ii) The graph of a second quadratic polynomial, $y = ax^2 + bx + c$, intersects the first graph at (−1, 7) and (3, 23). Find, in terms of a, the equation of the tangent to the second graph that is parallel to the chord joining the two points of intersection. [11]

(iii) The area enclosed between the two quadratic graphs is 472 square units. Find the coordinates of the point where the perpendicular bisector of the two points of intersection meets the tangent from part **(ii)**. [11]

BLANK PAGE

UNIVERSITY OF CAMBRIDGE INTERNATIONAL EXAMINATIONS
Cambridge International Level 3 Pre-U Certificate
Short Course

MATHEMATICS (STATISTICS WITH PURE MATHEMATICS) **1347/01**

Paper 1 Pure Mathematics **For Examination from 2012**

SPECIMEN MARK SCHEME

MAXIMUM MARK: 65

This document consists of **4** printed pages.

1	**(i)** $x(x+10) = 39 \Rightarrow x^2 + 10x = 39$ $\Rightarrow x^2 + 10x + 25 = 39 + 25$ $\Rightarrow (x+5)^2 = 64$		Or equivalent method	
	$a = 5$ and $b = 64$	B1 B1	$a = 5$ or $(x+5)^2$ seen $b = 64$ or expression $(x+a)^2 = 64$ seen	**2**
	(ii) Width = 3 metres Length = 13 metres	B1	3 and 13 cao	**1**
2	**(i)** $S = kE^n \Rightarrow \ln S = \ln(kE^n) = \ln k + \ln(E^n)$ $\Rightarrow \ln S = \ln k + n \ln E$	M1	Taking logs and using log laws to achieve $\ln S = \ln k + n \ln E$ or equivalent Log method must be seen for M mark	
	Compare with $Y = C + MX$ $\Rightarrow$ points should fit a straight line with gradient n (and intercept $\ln k$), hence $n = 0.0445$	A1	$n = 0.0445$ (given in question) Also accept $n = 0.044$ or $n = 0.045$	
	When $E = 1$, $S = k \Rightarrow k = 92$	B1	May use other points, $k = 92 \pm 0.5$	**3**
	(ii) $E = 15 \Rightarrow S = 92(15^{0.0445})$ $= 103.78$ dB	M1 A1	Putting $E = 15$ in formula with their k, n 103 to 104.5 from correct working	**2**
3	**(i)** $(2+3x)^5 = (2)^5 + 5(2)^4(3x) + 10(2)^3(3x)^2 + \ldots$ $= 32 + 240x + 720x^2 + \ldots$ Or $2^5(1+1.5x)^5 = 32\big(1 + 5(1.5)x + 10(1.5x)^2 + \ldots\big)$ $32(1 + 7.5x + 22.5x^2 + \ldots)$ $32 + 240x + 720x^2 + \ldots$	M1 B1 A1 A1	Coefficients 5 and 10 used correctly 32 as constant term or as common factor $240x$ $720x^2$	**4**
	(ii) Set $x = 0.01$ to get $(2.03)^5 \approx 32 + 2.4 + 0.072$ $= 34.472$ True value is 34.473 to 3dp, so value is correct to 2dp	M1 A1	Setting $x = 0.01$ to achieve 34.472 Showing that approximate value is correct to 2dp or stating that $(2.03)^5 = 34.47$ to 2dp	**2**
4	**(i)** $3x^2 - 2x + 2 = 4x + k \Rightarrow 3x^2 - 6x + (2-k) = 0$	B1	Equating expressions to form given quadratic	**1**
	(ii) $b^2 - 4ac = (-6)^2 - 4(3)(2-k)$ $= 12 + 12k$	B1	$12 + 12k$	
	Tangent cuts curve only once $\Rightarrow 12 + 12k = 0$ $\Rightarrow k = -1$	M1 A1	Solving discriminant = 0 $k = -1$ (cao)	
	Hence $3x^2 - 6x + 3 = 0 \Rightarrow 3(x-1)^2 = 0 \Rightarrow x = 1$ Tangent meets curve at (1, 3)	M1 A1	Substituting into quadratic, or equivalent (1, 3) (cao)	**5**

5	**(i)** $\sqrt{(2-1)^2+(9-2)^2} = \sqrt{50}$	B1	$\sqrt{50}$ or $5\sqrt{2}$	**1**
	(ii) Gradient of $AB = \frac{3}{4}$	M1	Either correct	
	Gradient of $BC = -\frac{4}{3}$	A1	Both correct	
	AB and BC are perpendicular	B1	Perpendicular, or equivalent	**3**
	(iii) The angle in a semi-circle is a right angle, so AC is a diameter of the circle			
	Centre is midpoint of AC	M1	Calculating midpoint of AC, or equivalent	
	i.e. $\left(1\frac{1}{2}, 5\frac{1}{2}\right)$	A1	$\left(1\frac{1}{2}, 5\frac{1}{2}\right)$	
	Radius $= \frac{5\sqrt{2}}{2}$ or $\frac{5}{\sqrt{2}}$	B1	A correct expression involving $\sqrt{2}$	**3**
6	**(i)** $y = -2x^4 + ax^2 + bx + 8 \Rightarrow \frac{dy}{dx} = -8x^3 + 2ax + b$	M1	Calculating $\frac{dy}{dx}$ correctly	
	$\frac{dy}{dx} = 0$ when $x = 0 \Rightarrow b = 0$	A1	$b = 0$ from completely correct method	
	$\frac{d^2y}{dx^2} = -24x^2 + 2a$	M1	Calculating $\frac{d^2y}{dx^2}$ correctly	
	$a > 0 \Rightarrow \frac{d^2y}{dx^2} > 0$ when $x = 0 \Rightarrow$ minimum t.p.	A1	Showing that $a > 0$ means that the t.p. is a minimum (result given in question)	**4**
	(ii) $y = -2x^4 + ax^2 + 8$			
	$\frac{dy}{dx} = 0 \Rightarrow -8x^3 + 2ax = 0$	M1	Solving their $\frac{dy}{dx} = 0$	
	$\Rightarrow x = 0$ or $4x^2 = a$	A1	$4x^2 = a$, or implied from $x = \frac{1}{2}\sqrt{a}$	
	So graph has stationary points when $x = \pm\frac{1}{2}\sqrt{a}$	A1	Both values given, from correct working	**3**
	(iii) To change the y values to m^2 he must divide the function by 10 000 and to change the x values to m he must replace x by $100x$ throughout	B1	Stating that using m^2 instead of cm^2 involves a factor of 10 000, or showing division by 100^2	
	$y = 0.0001\left(-2(100x)^4 + 6(100x)^2 + 8\right)$	B1	Stating that using m instead of cm involves a factor of 100, or showing replacing x by $100x$	**2**
	$= 0.0001(-200\,000\,000x^4 + 60\,000x^2 + 8)$			
	$= -20\,000x^4 + 6x^2 + 0.0008$			
	(iv) Volume $= \int_0^{0.02} (-20\,000x^4 + 6x^2 + 0.0008)\,dx$	M1	Correct integral, with or without limits, or equivalent using cm and cm^2	
	$= \left[-4000x^5 + 2x^3 + 0.0008x\right]_0^{0.02}$	A1	Integration correct for their valid expression	
	$= 0.000\,019\,2\ m^2$	M1	Substituting valid limits into their integral	
		A1	0. 000 019 2 (cao)	**4**

7	**(i)** *EITHER*: (geometry of completed square form)			
	$A = -1$	M1	$A = -1$ or (their) $P = 2$	
	$B = 7$	M1	$B = 7$ or (their) $Q = 8$	
	So $y = (x + 1)^2 + 7$ and $x = 3 \Rightarrow y = 23$	A1	Showing that $y = 23$ (given)	
	OR: (calculus with completed square form)		Accept alternative equivalent methods and mark accordingly	
	$y = (x - A)^2 + B \Rightarrow \frac{dy}{dx} = 2(x - A)$			
	$\frac{dy}{dx} = 0$ when $x = -1 \Rightarrow A = -1$			
	$y = 7$ when $x = -1 \Rightarrow B = 7$			
	So $y = (x + 1)^2 + 7$ and $x = 3 \Rightarrow y = 23$			
	OR: (calculus with expanded form)			
	$y = x^2 + Px + Q \Rightarrow \frac{dy}{dx} = 2x + P$			
	$\frac{dy}{dx} = 0$ when $x = -1 \Rightarrow P = 2$			
	$y = 7$ when $x = -1 \Rightarrow Q = 8$			
	So $y = x^2 + 2x + 8$ and $x = 3 \Rightarrow y = 23$			**3**
	(ii) $y = ax^2 + bx + c$			
	$(-1, 7)$ is on graph $\Rightarrow a - b + c = 7$	M1	$a - b + c = 7$	
	$(3, 23)$ is on graph $\Rightarrow 9a + 3b + c = 23$	M1	$9a + 3b + c = 23$	
	Hence $8a + 4b = 16 \Rightarrow b = 4 - 2a$	A1	$b = 4 - 2a$	
	$\Rightarrow c = 11 - 3a$	A1	$c = 11 - 3a$	
			Or implied by subsequent working	
	$\frac{dy}{dx} = 2ax + b$ or $2ax + 4 - 2a$	M1	$\frac{dy}{dx} = 2ax + b$ in any valid form	
	Gradient of chord joining $(-1, 7)$ to $(3, 23)$ is 4	B1	Gradient $= (16 - 7) \div (3 - -1)$	
	$\frac{dy}{dx} = 4 \Rightarrow x = \frac{4 - b}{2a} \Rightarrow x = 1$	A1	$x = 1$ (cao)	
	$y = ax^2 + (4 - 2a)x + (1 - 3a)$	M1	y in terms of a and x using their b and c	
	$x = 1 \Rightarrow y = 15 - 4a$	A1	Substituting $x =$ (their) 1 into (their) y	
	So equation of tangent is $y - 15 + 4a = 4(x - 1)$	M1	Hence, (their) tangent equation	
	$\Rightarrow y = 4x + 11 - 4a$	A1	Correct expression (cao)	**11**
			Accept alternative equation methods and mark accordingly	
	(iii) $472 = \int_{-1}^{3} (x^2 + 2x + 8)\,dx -$	B1	(their) $\int_{-1}^{3} (x^2 + 2x + 8)\,dx$	
	$\left[a\int_{-1}^{3} (x^2 - 2x - 3)\,dx + \int_{-1}^{3} (4x + 11)\,dx\right]$	B1	(their) $a\int_{-1}^{3} (x^2 - 2x - 3)\,dx + \int_{-1}^{3} (4x + 11)\,dx$	
	$\Rightarrow 472 = \left[\frac{1}{3}x^3 + x^2 + 8x\right]_{-1}^{3} -$			
	$a\left[\frac{1}{3}x^3 - x^2 - 3x\right]_{-1}^{3} - \left[2x^2 + 11x\right]_{-1}^{3}$	M1	Setting difference = 472 and simplifying	
	$\Rightarrow 472 = 49\frac{1}{3} - a(-10\frac{2}{3}) - 60$			
	$\Rightarrow 472 = 109\frac{1}{3} + 10\frac{2}{3}a \Rightarrow a = 34$	A1	(their) a (correct to 3sf)	
	Tangent is $y = 4x - 125$	B1	$y = 4x + 11 -$ (their) $4a$	
	Perpendicular bisector has gradient $-\frac{1}{4}$	M1	$-1 \div$ (their) tgt gradient, may be implied	
	and passes through $(1, 15)$	M1	Midpt $\left(1, \frac{1}{2}(7 + \text{(their)}23)\right)$, may be implied	
	So perpendicular bisector is $y = -\frac{1}{4}x + 15\frac{1}{4}$	A1	$y = -\frac{1}{4}x + 15\frac{1}{4}$ (cao)	
	The lines meet when $4x - 125 = -\frac{1}{4}x + 15\frac{1}{4}$	M1	Solving where (their) lines meet	
	$\Rightarrow 17x = 561 \Rightarrow x = 33$	A1	(their) x-value (or y-value if y found first)	
	Lines meet at $(33, 7)$	B1	$(33, 7)$ (cao)	**11**

UNIVERSITY OF CAMBRIDGE INTERNATIONAL EXAMINATIONS
Cambridge International Level 3 Pre-U Certificate
Short Course

MATHEMATICS (STATISTICS WITH PURE MATHEMATICS) **1347/02**

Paper 2 Statistics **For Examination from 2012**

SPECIMEN PAPER

2 hours

Additional Materials: Answer Booklet/Paper
Graph Paper
List of Formulae (MF21)

READ THESE INSTRUCTIONS FIRST

If you have been given an Answer Booklet, follow the instructions on the front cover of the Booklet.
Write your Centre number, candidate number and name on all the work you hand in.
Write in dark blue or black pen.
You may use a soft pencil for any diagrams or graphs.
Do not use staples, paper clips, highlighters, glue or correction fluid.

Answer **all** the questions.
Give non-exact numerical answers correct to 3 significant figures, or 1 decimal place in the case of angles in degrees, unless a different level of accuracy is specified in the question.
The use of an electronic calculator is expected, where appropriate.
You are reminded of the need for clear presentation in your answers.

At the end of the examination, fasten all your work securely together.
The number of marks is given in brackets [ ] at the end of each question or part question.
The total number of marks for this paper is 80.

This document consists of **6** printed pages.

UNIVERSITY of CAMBRIDGE International Examinations

[Turn over

1 The table shows information about the twenty most densely populated countries in the world.

Country	Population	Rank	Land Area (km^2)	Rank	Population density
Monaco	32 410	16	2	19	16 205
Singapore	4 425 700	8	693	10	6 386
Malta	398 500	12	316	13	1 261
Maldives	349 100	13	300	14	1 164
Bahrain	688 300	11	665	11	1 035
Bangladesh	144 319 600	2	144 000	3	1 002
Vatican City	920	20	1	20	920
Barbados	279 300	14	431	12	648
Nauru	13 050	18	21	18	621
Mauritius	1 230 600	10	2 040	9	603
South Korea	48 422 600	4	98 480	4	492
San Marino	28 800	17	61	16	472
Tuvalu	11 640	19	26	17	448
Netherlands	16 407 500	5	41 526	5	395
Lebanon	3 826 000	9	10 400	8	368
Belgium	10 364 400	6	30 510	6	340
Japan	127 417 200	3	377 835	2	337
India	1 080 264 400	1	3 287 590	1	329
Marshall Islands	59 070	15	181	15	326
Rwanda	8 440 800	7	26 338	7	320

source: http://www.worldatlas.com

(i) Write down the median and quartiles for the populations of these twenty countries. What do these values tell you about the shape of the distribution of the populations? Identify the countries for which the population is an outlier for this sample. [6]

(ii) Explain why it is not appropriate to use the product-moment correlation coefficient as a measure of correlation between the populations and the areas for these twenty countries. Show that the value of Spearman's rank correlation coefficient between the populations and the areas is 0.9789. [4]

2 Louise has made up two sets of words for a Psychology project. Each set consists of twenty words written in five groups of four. On card *A* the words in each group are random; on card *B* the words in each group have a common theme.

Card *A*

Rabbit	Turnip	Daisy	Lamp
Scarf	Pebble	Cheese	Pencil
Snow	Triangle	Yacht	Bag
Green	Pram	Chair	Film
Tennis	Bottle	String	Needle

Card *B*

Rabbit	Dog	Cat	Mouse
Scarf	Hat	Coat	Gloves
Snow	Rain	Wind	Sun
Green	Blue	Red	Yellow
Tennis	Golf	Hockey	Squash

Louise asks six volunteers to try to memorise the words on card *A* and another six to try to memorise the words on card *B*. Each group is given 30 seconds and is then asked to write down as many of the words as they can remember.

The number of words that each volunteer remembered correctly is given below.

Card *A*:	4	9	3	14	7	11
Card *B*:	12	18	13	15	5	8

(i) Carry out a Wilcoxon rank-sum test, at the 5% level, to test whether card *B* was easier to remember than card *A*. [4]

Suppose that, instead, Louise had asked six volunteers to try to memorise the words on card *A* and had then asked the same six to try to memorise the words on card *B*. Assume that the results above represent the paired data showing the number of words remembered correctly from each list by each of these six volunteers.

(ii) Carry out a Wilcoxon matched-pairs signed-rank test, at the 5% level, to test whether card *B* was easier to remember than card *A*. [4]

(iii) Apart from changing the words on the lists, give one way in which Louise could have improved the design of her experiment in part **(i)**, and a different way in which she could have improved the design of the experiment in part **(ii)**. [2]

3 An article from the organic coffee producers association claimed that 5% of the coffee bought in the UK is made with beans from fair-trade sources.

A high-street coffee vendor sells different types of coffee, some of which are made with beans from fair-trade sources. One day he sells 10 cups of coffee in the first half hour of trading. He uses a binomial B(n, p) distribution to model the number of these cups that are made with beans from fair-trade sources.

(i) Calculate the probability that at least 2 of the 10 cups are made with beans from fair-trade sources. State the assumptions you have made. [4]

The vendor finds that exactly 1 of the 10 cups was made with beans from fair-trade sources. He claims that this shows that more than 5% of the coffee bought in his shop is made with beans from fair-trade sources.

(ii) Test whether this result supports the vendor's claim, assuming that a binomial model is valid. Use a 20% level of significance for your test. [4]

Each day for a month the vendor records how many of the first ten cups of coffee sold were made with beans from fair-trade sources. His results are given below.

Number of cups made with beans from fair-trade sources	Number of days
0	4
1	7
2	9
3	6
4	3
5	1
6 or more	0

(iii) Calculate the mean for this sample and hence estimate an appropriate value for the percentage of the vendor's coffee sales that are made with beans from fair-trade sources. [3]

(iv) Use the vendor's records to calculate an estimate for the population variance. Compare this with the theoretical value from B(n, p) using your estimate for p. [4]

(v) Suggest two improvements to the vendor's sampling method. [2]

4 An economist records the daily rises and falls of the £ sterling against the US dollar. At the start of each business day she records the value of £1 in US dollars and then calculates the increase or decrease from the previous day in \$0.001.

She claims that the changes can be modelled using a normal distribution N(0, 100).

(i) Assuming the claim to be true, calculate the probability of a change that is

(a) greater than 16, [3]

(b) more negative than −0.7. [3]

Each working week there are five daily changes.

(ii) Assuming the economist's claim to be true, write down the distribution for the mean of the weekly changes and calculate the probability of a mean change of between ±7.8. [4]

One week the economist finds that the five changes are 6, 16, −12, 2, −7.

She puts these values into a spreadsheet and calculates various totals.

	A	B	C	D	E
1	6	36	5	25	
2	16	256	15	225	
3	−12	144	−13	169	
4	2	4	1	1	
5	−7	49	−8	64	
6					
7	5	489	0	484	

(iii) Use this result to determine a 95% confidence interval for the mean change, assuming that the changes can be regarded as being a random sample from a normal population. [6]

The economist then collects data for 100 daily changes. She finds that 25 of the 100 changes were greater than 9.2 and just one was greater than 29.

(iv) Assuming that a normal model is appropriate and that the data are typical, calculate appropriate values (to 2 significant figures) for the mean and standard deviation of the daily changes. [5]

5 A biologist carries out a study into the effect of different chemicals on plant growth. He grows 50 common vetch plants in ordinary conditions, 25 in an environment where the air is polluted with motor exhaust fumes and 25 in soil polluted with oil. He records what happens to the leaves of the plants. The frequencies are given in the table.

	Grew normally	Spotted leaves	Plant died	TOTAL
Control	31	12	7	50
Polluted air	12	10	3	25
Polluted soil	7	8	10	25
TOTAL	50	30	20	100

The biologist suspects that there is an association between the type of pollution and what happens to the leaves of the plants. A χ^2 test is to be applied to the data in the contingency table.

(i) State the null and alternative hypotheses for the test. [1]

(ii) Assuming the null hypothesis to be true, how many of the plants grown in polluted soil would you expect to have grown normally? Show that the contribution to the χ^2 test statistic from the cell for plants grown in polluted soil that grew normally is 2.42 to 2 decimal places. [2]

(iii) Carry out the test at the 5% level of significance. State the conclusion that the biologist should make. [8]

The biologist then measures the heights, in cm, of the 50 plants that grew normally. The heights of the 31 plants in the control group can be modelled using a normal distribution with mean 35 and variance 9. The heights of the 12 plants grown in polluted air can also be modelled using a normal distribution with variance 9. The heights of the 7 plants grown in polluted soil cannot be modelled using a normal distribution.

(iv) The sample mean height for the 12 plants grown in polluted air was 32 cm. Test, at the 5% level, whether the average height of plants grown in polluted air is significantly less than 35 cm. [6]

It is later found that the heights of common vetch plants grown in polluted air can be modelled using a normal distribution with mean 31.5 and variance 9.

(v) Calculate the probability of having made a Type I error and of having made a Type II error in your test from part **(iv)**. [3]

(vi) Using the data for the 7 plants grown in polluted soil, what test could the biologist use to test whether the average height of all common vetch plants grown in polluted soil is less than 35 cm? [2]

UNIVERSITY OF CAMBRIDGE INTERNATIONAL EXAMINATIONS
Cambridge International Level 3 Pre-U Certificate
Short Course

MATHEMATICS (STATISTICS WITH PURE MATHEMATICS) **1347/02**

Paper 2 Statistics **For Examination from 2012**

SPECIMEN MARK SCHEME

MAXIMUM MARK: 80

This document consists of **6** printed pages.

UNIVERSITY of CAMBRIDGE
International Examinations

[Turn over

1	**(i)** By population, 10th is 1 230 600 and 11th is 688 300 so median is 959 450	B1	959 450 cao	
	5th is 16 407 500 and 6th is 10 364 400 so upper quartile is 13 385 950	B1	Any value in $10\,364\,400 < Q_3 < 16\,407\,500$ (need not be identified as upper quartile)	
	15th is 59 070 and 16th is 32 410 so lower quartile is 45 740		Any value in $32\,410 < Q_1 < 59\,070$ (need not be identified as lower quartile)	
	Distribution would appear to have positive skew	B1		
	IQR = 13 385 950 − 45 740 = 13 340 210	M1	Calculating IQR from their reasonable quartiles	
	Outliers are values outside $Q_2 \pm 1.5 \times \text{IQR}$ 959 450 − 20 010 315 = −19 050 865 < 0 959 450 + 20 010 315 = 20 969 765	A1	Calculating their median + 1.5 × their IQR	
	Hence: India, Bangladesh, Japan, South Korea	B1	India, Bangladesh, Japan, South Korea	**6**
	(ii) Not random samples, distributions are skewed so unlikely to be from normal distributions	B1	Not normal, with a supporting reason	
	16 8 12 13 11 2 20 14 18 10 4 17 19 5 9 6 3 1 15 7 19 10 13 14 11 3 20 12 18 9 4 16 17 5 8 6 2 1 15 7 −3 −2 −1 −1 0 1 0 2 0 1 0 1 2 0 1 0 1 0 0 0			
	$\Sigma d^2 = 28$	M1	Substantially correct calculation of Σd^2	
	$r_s = 1 - \dfrac{6 \times 28}{20(400-1)} = 1 - 0.021\,05$	M1	Correct method for 0.021 05	
	= 0.9789 (4 dp) AG	A1	Achieving given value from correct working	**4**

2	**(i)** H_0 : samples come from the same population H_1 : A tends to have lower ranks than B One-sided $\alpha = 5\%$			
	Rank: 1 2 3 4 5 6 7 8 9 10 11 12 3 4 5 7 8 9 11 12 13 14 15 18 *A A B A B A A B B A B B*	M1	Substantially correct method (may rank in reverse order)	
	Rank sum for $A = 30$ $m = 6$ Rank sum for $B = 48$ $n = 6$ $W = 30$	A1	$W = 30$ from correct working	
	Critical value for W is 28	B1	Critical value = 28	
	Accept H_0 Insufficient evidence at the 5% level to be able to conclude that list B is easier to remember than list A	B1	Correct conclusion, in context	**4**
	(ii) H_0 : $B - A$ differences symmetric about 0 H_1 : $B - A$ differences tend to be positive One-sided $\alpha = 5\%$			
	Subject: 1 2 3 4 5 6 A 4 9 3 14 7 11 B 12 18 13 15 5 8 $B - A$ 8 9 10 1 −2 −3			
	Rank: 1 2 3 4 5 6 $\lvert B - A \rvert$ 1 2 3 8 9 10 sign + − − + + +	M1	Substantially correct method	
	$P = 16$ $Q = 5$ $\Rightarrow T = 5$	A1	$T = 5$ from correct working	
	Critical value for T is 2	B1	Critical value = 2	
	Accept H_0 Insufficient evidence at the 5% level to be able to conclude that list B is easier to remember than list A	B1	Correct conclusion, in context	**4**
	(iii) In **(i)** she could improve the experiment by increasing the sample sizes (asking more people)	B1	Increase sample size	
	In **(ii)** she could improve the experiment by randomising which card people looked at first	B1	Randomise order	**2**

3	**(i)** X = number of cups made with beans from fair-trade sources $X \sim B(10, 0.05)$			
	$P(X \geqslant 2) = 1 - P(X \leqslant 1) = 1 - 0.9139$ $= 0.0861$	M1 A1	$P(X \leqslant 1) = 0.9139$ 0.0861 from correct working	
	Assume that purchases are independent and that probability of choosing coffee made with beans from fair-trade sources is 5%	B1 B1	Independence (not sufficient just to say that we are assuming a binomial model is valid) $p = 0.05$	**4**
	(ii) $X \sim B(10, p)$ $H_0 : p = 0.05$ $\alpha = 20\%$ $H_1 : p > 0.05$ one-sided test	B1	Null and alternative hypotheses correct	
	Assuming H_0: $X \sim B(10, 0.05)$ Expected (mean) $X = 0.5$ Observed value of X is 1 $P(X \geqslant 1) = 1 - 0.5987 = 0.4013$	B1	$P(X \geqslant 1) = 0.4013$	
	$X = 1$ is not (entirely) in the tail 20% Accept H_0 At the 80% level of confidence we have insufficient evidence to support the vendor's claim The data are not inconsistent with the probability being 5%	M1 A1	Accept H_0 (even if working incorrect) A correct conclusion stated in words Insufficient evidence to support vendor's claim	**4**
	(iii)			

x	0	1	2	3	4	5	>5
f	4	7	9	6	3	1	0

	$\Sigma f = 30$ $\Sigma fx = 60$	M1	$\Sigma f = 30$ and $\Sigma fx = 60$
	$\bar{x} = \frac{60}{30} = 2$	A1	2
	Mean is 2 cups out of 10, so estimate of p is 0.2 or 20%	B1	Their mean divided by 10 (decimal or %) **3**
	(iv) $\Sigma f(x - \bar{x})^2 = 50$	M1	Substantially correct working
	$s^2 = 50 \div 29 = 1.724$	A1	1.72 or better
	$npq = 10 \times 0.2 \times 0.8 = 1.6$	M1	Calculating npq with $n = 10$ and their p
	The estimated population variance is similar to the theoretical value, but is slightly larger	A1	Comparing their 1.72 with correct 1.6 **4**
	(v) Vendor should have used a larger sample	B1	Increasing sample size
	and should have chosen coffee sales randomly through the day	B1	Attempting to introduce independence or randomness **2**

4	**(i)** X is daily change (in \$0.001), where $X \sim \text{N}(0, 100)$			
	(a) $\text{P}(X > 16) = \text{P}(Z > 1.6)$ where $Z \sim \text{N}(0, 1)$	M1	1.6 seen or implied	
	$= 1 - \Phi(1.6) = 1 - 0.9452$	M1	0.9452 or their $\Phi(z)$	
	$= 0.0548$	A1	0.0548 cao	**3**
	(b) $\text{P}(X < -0.7) = \text{P}(X > 0.7) = \text{P}(Z > 0.07)$	M1	± 0.07 seen or implied	
	$= 1 - \Phi(0.07) = 1 - 0.5279$	M1	0.5279 or their $\Phi(z)$	
	$= 0.4721$	A1	0.4721 cao	**3**
	(ii) $\overline{X} \sim \text{N}(0, 20)$	B1	cao	
	$\text{P}(-7.8 < \overline{X} < 7.8) = \text{P}(-1.744 < Z < 1.744)$	M1	1.744 seen or implied	
	$= 2\left(\Phi(1.744) - 0.5\right)$			
	$= 2 \times 0.4595$	M1	Valid method for their z	
	$= 0.9190$	A1	0.9190 cao	**4**
	(iii) For the sample data: $n = 5$, $\Sigma x = 5 \Rightarrow \bar{x} = 1$	B1	Mean = 1	
	$\Sigma\left(x - \bar{x}\right)^2 = 484 \Rightarrow s^2 = 121$	M1	$484 \div (5 - 1)$	
	$\Rightarrow s = 11$	A1	11 cao	
	For a 95% CI, critical values of t_4 are ± 2.776	B1	2.776 seen	
	Confidence limits are $1 \pm 2.776 \times \dfrac{11}{\sqrt{5}}$	M1	Correct method for CI	
	$= 1 \pm 2.776 \times 4.919$			
	$= 1 \pm 13.656$			
	Confidence interval is $(-12.66, 14.66)$	A1	cao	**6**
	(iv) X is daily change (in \$0.001), where $X \sim \text{N}(\mu, \sigma^2)$			
	$\text{P}(X > 29) = 0.01 \Rightarrow \dfrac{29 - \mu}{\sigma} = 2.326$	M1	A correct expression using 2.326	
	$\text{P}(X > 9.2) = 0.25 \Rightarrow \dfrac{9.2 - \mu}{\sigma} = 0.674$	M1	A correct expression using 0.674	
	Hence $29 - \mu = 2.326\sigma$ and $9.2 - \mu = 0.674\sigma$			
	$\Rightarrow 1.652\sigma = 19.8$	M1	Correct method from their reasonable equations, or equivalent	
	$\Rightarrow \sigma = 12$ and $\mu = 1.1$ to 2sf	A1	12 cao	
		A1	1.1 or 1.2	**5**

5 **(i)** H_0: Independence (no association) H_1: Dependence exists between the type of pollution and what happens to the leaves	B1	Both hypotheses correct	**1**
(ii) $0.50 \times 0.25 \times 100 = 12.5$	B1	12.5	
$(7 - 12.5)^2 \div 12.5 = 2.42$ (answer given)	B1	(7 – their 12.5)^2 ÷ (their 12.5)	**2**
(iii) Obs: 31 12 7 / 12 10 3 / 7 8 10 Exp: 25 15 10 / 12.5 7.5 5 / 12.5 7.5 5	B1	Expected frequencies correct (seen or implied)	
χ^2: 1.44 0.60 0.90 / 0.02 0.83 0.80 / 2.42 0.03 5.00 = 12.04(6)	M1 A1	Calculating χ^2 correctly 12.04 to 12.05	
$\nu = (3 - 1) \times (3 - 1) = 4$	M1	4 degrees of freedom	
Critical value = 9.488	A1	9.488	
12.046 > critical value ⇒ reject H_0	M1	Correct conclusion for their values	
There seems to be dependence between the type of pollution and what happens to the leaves	A1	Their conclusion in context	
Specifically, far more of the plants grown in polluted soil die than would be expected if there were independence	A1	Specifically identifying where the greatest contribution to χ^2 comes from, in context	**8**
(iv) X is height of plants grown in polluted air, where $X \sim N(\mu, 9)$			
$H_0 : \mu = 35$ $\alpha = 5\%$	B1	Null hypothesis correct	
$H_1 : \mu < 35$ one-sided	B1	Correct one-sided alternative hypothesis	
Assuming H_0, $\overline{X} \sim N(35, 0.75)$			
EITHER: $P(\overline{X} < 32) = P(Z < -3.464) < 0.05$	M1	Appropriate calculation	
OR: Critical value = $35 - 1.645\sqrt{0.75} = 33.57$	A1	–3.464 or 33.57	
Reject H_0	M1	Reject H_0 from correct calculations	
Data support claim that mean height for plants grown in polluted air is less than 35 cm	A1	Correct interpretation from correct calculations	**6**
(v) Critical value = 33.57			
P(Type I error) = P(falsely reject H_0) = $P(\overline{X} < 33.57)$ when $\overline{X} \sim N(35, 0.75) = 0.05$	B1	P(Type I) = 0.05 or 5%	
P(Type II error) = P(falsely accept H_0)			
$= P(\overline{X} > 33.57)$ when $\overline{X} \sim N(31.5, 0.75)$	M1	Correct calculation	
$= P(Z > 2.40) = 1 - 0.9918 = 0.0082$	A1	0.0082 or 0.82%	**3**
(vi) Sign test or Wilcoxon signed rank test to compare median with 35	M1 A1	Any appropriate non-parametric test Median = 35	**2**