

Topic 4 Part 1 [219 marks]

Minta surveyed students from her school about their preferred morning snack from a choice of an apple, a fruit salad or a smoothie.

She surveyed 350 students, of whom 210 are female.

She performed a χ^2 test at the 5% significance level to determine whether there is a relationship between the choice of morning snack and gender.

1a. State Minta's null hypothesis.

[1 mark]

.....
.....
.....

1b. State the number of degrees of freedom.

[1 mark]

.....
.....
.....

1c. 150 students showed a preference for a smoothie.

[2 marks]

Calculate the expected number of female students who chose a smoothie.

.....
.....
.....

- 1d. Minta found that the calculated value of the χ^2 test was 3.576. The critical value at the 5% significance level is 5.99. [2 marks]

State Minta's conclusion. Give a reason for your answer.

.....

.....

.....

The daily January temperature of Cairns is normally distributed with a mean of 34°C and a standard deviation of 3.

- 2a. Calculate the probability that the temperature on a randomly chosen day in January is less than 39°C. [2 marks]

.....

.....

.....

- 2b. Calculate the expected number of days in January that the temperature will be more than 39°C. [2 marks]

.....

.....

.....

2c. On a randomly chosen day in January, the probability that the temperature is above T °C is 0.7.

[2 marks]

Find the value of T .

.....

.....

.....

Identical mosquito traps are placed at different distances from a lake. On one day the number of mosquitoes caught in 10 of the traps is recorded.

Distance, m (x)	8	15	22	30	34	45	50	60	74	82
Number of mosquitoes (y)	78	75	72	67	66	59	59	53	48	43

It is believed the number of mosquitoes caught varies linearly with the distance, in metres, of the trap from the lake.

3a. Find

[4 marks]

- (i) Pearson's product-moment correlation coefficient, r ;
- (ii) the equation of the regression line y on x .

.....

.....

.....

3b. Use the equation of the regression line y on x to estimate the number of mosquitoes caught in a trap that is 28 m from the lake. [2 marks]

.....

.....

.....

The weight, W , of bags of rice follows a normal distribution with mean 1000 g and standard deviation 4 g.

- 4a. Find the probability that a bag of rice chosen at random weighs between 990 g and 1004 g. [2 marks]

.....

.....

.....

- 4b. 95% of the bags of rice weigh less than k grams. [2 marks]

Find the value of k .

.....

.....

.....

- 4c. For a bag of rice chosen at random, $P(1000 - a < W < 1000 + a) = 0.9$. [2 marks]

Find the value of a .

.....

.....

.....

The producer of a TV dancing show asked a group of 150 viewers their age and the type of Latin dance they preferred. The types of Latin dances in the show were Argentine tango, Samba, Rumba and Cha-cha-cha. The data obtained were organized in the following table.

	Dance			
	Argentine tango	Samba	Rumba	Cha-cha-cha
20 years old and younger	35	23	12	10
Older than 20 years old	20	17	18	15

A χ^2 test was carried out, at the 5% significance level.

5a. Write down the null hypothesis for this test.

[1 mark]

.....

.....

.....

5b. Write down the observed number of viewers who preferred Rumba **and** were older than 20 years old.

[1 mark]

.....

.....

.....

5c. Use your graphic display calculator to find the p -value for this test.

[2 marks]

.....

.....

.....

5d. The producer claims that the type of Latin dance a viewer preferred is independent of their age.

[2 marks]

Decide whether this claim is justified. Give a reason for your decision.

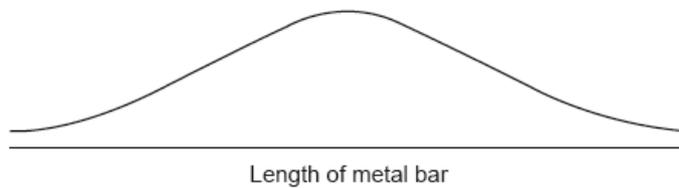
.....

.....

.....

A factory makes metal bars. Their lengths are assumed to be normally distributed with a mean of 180 cm and a standard deviation of 5 cm.

6a. On the following diagram, shade the region representing the probability that a metal bar, chosen at random, will have a length less than 175 cm. [2 marks]



.....

.....

.....

6b. A metal bar is chosen at random.

[4 marks]

(i) The probability that the length of the metal bar is less than 175 cm is equal to the probability that the length is greater than h cm. Write down the value of h .

(ii) Find the probability that the length of the metal bar is greater than one standard deviation above the mean.

.....

.....

.....

In a debate on voting, a survey was conducted. The survey asked people's opinion on whether or not the minimum voting age should be reduced to 16 years of age. The results are shown as follows.

	Age 18–25	Age 26–40	Age 41+	Total
Oppose the reduction	12	20	48	80
Favour the reduction	18	15	17	50
Total	30	35	65	130

A χ^2 test at the 1% significance level was conducted. The χ^2 critical value of the test is 9.21.

7a. State

[2 marks]

- (i) H_0 , the null hypothesis for the test;
- (ii) H_1 , the alternative hypothesis for the test.

.....

.....

.....

7b. Write down the number of degrees of freedom.

[1 mark]

.....

.....

.....

7c. Show that the expected frequency of those between the ages of 26 and 40 who oppose the reduction in the voting age is 21.5, correct to three significant figures.

[2 marks]

.....

.....

.....

7d. Find

[3 marks]

- (i) the χ^2 statistic;
- (ii) the associated p -value for the test.

.....

.....

.....

7e. Determine, giving a reason, whether H_0 should be accepted.

[2 marks]

.....

.....

.....

Daniel grows apples and chooses at random a sample of 100 apples from his harvest.

He measures the diameters of the apples to the nearest cm. The following table shows the distribution of the diameters.

Diameter (to the nearest cm)	5	6	7	8	9
Frequency	15	27	33	17	8

8a. Using your graphic display calculator, write down the value of

[3 marks]

- (i) the mean of the diameters in this sample;
- (ii) the standard deviation of the diameters in this sample.

.....

.....

.....

- 8b. Daniel assumes that the diameters of all of the apples from his harvest are normally distributed with a mean of 7 [3 marks] cm and a standard deviation of 1.2 cm. He classifies the apples according to their diameters as shown in the following table.

Classification	Diameter (cm)
Small	Diameter < 6.5
Medium	$6.5 \leq \text{Diameter} < a$
Large	Diameter $\geq a$

Calculate the percentage of **small** apples in Daniel's harvest.

.....

.....

.....

- 8c. Daniel assumes that the diameters of all of the apples from his harvest are normally distributed with a mean of 7 [2 marks] cm and a standard deviation of 1.2 cm. He classifies the apples according to their diameters as shown in the following table.

Classification	Diameter (cm)
Small	Diameter < 6.5
Medium	$6.5 \leq \text{Diameter} < a$
Large	Diameter $\geq a$

Of the apples harvested, 5% are **large** apples.

Find the value of a .

.....

.....

.....

- 8d. Daniel assumes that the diameters of all of the apples from his harvest are normally distributed with a mean of 7 [2 marks] cm and a standard deviation of 1.2 cm. He classifies the apples according to their diameters as shown in the following table.

Classification	Diameter (cm)
Small	Diameter < 6.5
Medium	$6.5 \leq \text{Diameter} < a$
Large	Diameter $\geq a$

Find the percentage of **medium** apples.

.....

.....

.....

- 8e. Daniel assumes that the diameters of all of the apples from his harvest are normally distributed with a mean of 7 [2 marks] cm and a standard deviation of 1.2 cm. He classifies the apples according to their diameters as shown in the following table.

Classification	Diameter (cm)
Small	Diameter < 6.5
Medium	$6.5 \leq \text{Diameter} < a$
Large	Diameter $\geq a$

This year, Daniel estimates that he will grow 100 000 apples.

Estimate the number of **large** apples that Daniel will grow this year.

.....

.....

.....

The following table shows the number of bicycles, x , produced daily by a factory and their total production cost, y , in US dollars (USD). The table shows data recorded over seven days.

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
Number of bicycles, x	12	15	14	17	20	18	21
Production cost, y	3900	4600	4100	5300	6000	5400	6000

- 9a. (i) Write down the Pearson's product-moment correlation coefficient, r , for these data. [4 marks]
(ii) Hence comment on the result.

.....

.....

.....

- 9b. Write down the equation of the regression line y on x for these data, in the form $y = ax + b$. [2 marks]

.....

.....

.....

- 9c. Estimate the total cost, **to the nearest USD**, of producing 13 bicycles on a particular day. [3 marks]

.....

.....

.....

9d. All the bicycles that are produced are sold. The bicycles are sold for 304 USD **each**. [2 marks]

Explain why the factory does **not** make a profit when producing 13 bicycles on a particular day.

.....

.....

.....

9e. All the bicycles that are produced are sold. The bicycles are sold for 304 USD **each**. [5 marks]

- (i) Write down an expression for the total selling price of x bicycles.
- (ii) Write down an expression for the **profit** the factory makes when producing x bicycles on a particular day.
- (iii) Find the least number of bicycles that the factory should produce, on a particular day, in order to make a profit.

.....

.....

.....

A group of 100 students gave the following responses to the question of how they get to school.

	Walk	Public transport	Car	Bicycle	Total
Female	18	13	14	3	48
Male	9	17	10	16	52
Total	27	30	24	19	100

A χ^2 test for independence was conducted at the 5% significance level. The null hypothesis was defined as

H_0 : Method of getting to school is independent of gender.

10a. Find the expected frequency for the females who use public transport to get to school. [2 marks]

10b. Find the χ^2 statistic. [2 marks]

10c. The

[2 marks]

χ^2 critical value is
7.815 at the
5% significance level.

State whether or not the null hypothesis is accepted. Give a reason for your answer.

A survey investigated the relationship between the number of cleaners,
 n , and the amount of time,
 t , it takes them to clean a school.

Number of cleaners, n	Time, t (minutes)
1	193
2	172
3	118
5	112
6	87

11a. Use your graphic display calculator to write down the equation of the regression line

[2 marks]

t on
 n .

11b. Write down the value of the Pearson's product-moment correlation coefficient,

[2 marks]

r .

11c. Use your regression equation to find the amount of time 4 cleaners take to clean the school.

[2 marks]

The heights of apple trees in an orchard are normally distributed with a mean of
3.42 m and a standard deviation of
0.21 m.

12a. Write down the probability that a randomly chosen tree has a height greater than

[1 mark]

3.42 m.

12b. Write down the probability that a randomly chosen tree will be within 2 standard deviations of the mean of

[1 mark]

3.42 m.

12c. Use your graphic display calculator to calculate the probability that a randomly chosen tree will have a height greater than

[2 marks]

3.35 m.

12d. The probability that a particular tree is less than

[2 marks]

x metres high is

0.65. Find the value of

x .

A study was carried out to determine whether the country chosen by students for their university studies was influenced by a person's gender. A random sample was taken. The results are shown in the following table.

	Country Chosen		
	USA	Australia	UK
Male	55	26	40
Female	25	31	41

A

χ^2 test was performed at the 1% significance level.

The critical value for this test is 9.210.

13a. State the null hypothesis.

[1 mark]

13b. Write down the number of degrees of freedom.

[1 mark]

13c. Write down

[2 marks]

- (i) the χ^2 statistic;
- (ii) the associated p -value.

13d. State, giving a reason, whether the null hypothesis should be accepted.

[2 marks]

As part of his IB Biology field work, Barry was asked to measure the circumference of trees, in centimetres, that were growing at different distances, in metres, from a river bank. His results are summarized in the following table.

Distance, x (metres)	5	12	17	21	24	30	34	44	47
Circumference, y (centimetres)	82	76	70	68	67	60	62	50	50

14a. State whether *distance from the river bank* is a continuous **or** discrete variable.

[1 mark]

14b. **On graph paper**, draw a scatter diagram to show Barry's results. Use a scale of 1 cm to represent 5 m on the x -axis and 1 cm to represent 10 cm on the y -axis. [4 marks]

14c. Write down

[2 marks]

- (i) the mean distance, \bar{x} , of the trees from the river bank;
- (ii) the mean circumference, \bar{y} , of the trees.

14d. Plot and label the point

[2 marks]

$M(\bar{x}, \bar{y})$ on your graph.

14e. Write down [4 marks]
(i) the Pearson's product-moment correlation coefficient,
 r , for Barry's results;
(ii) the equation of the regression line
 y on
 x , for Barry's results.

14f. Draw the regression line [2 marks]
 y on
 x on your graph.

14g. **Use the equation of the regression line** [2 marks]
 y on
 x to estimate the circumference of a tree that is 40 m from the river bank.

A group of candidates sat a Chemistry examination and a Physics examination. The candidates' marks in the Chemistry examination are normally distributed with a mean of 60 and a standard deviation of 12.

15a. Draw a diagram that shows this information. [2 marks]

15b. Write down the probability that a randomly chosen candidate who sat the Chemistry examination scored at most 60 marks. [1 mark]

15c. Hee Jin scored 80 marks in the Chemistry examination. [2 marks]
Find the probability that a randomly chosen candidate who sat the Chemistry examination scored **more** than Hee Jin.

15d. The candidates' marks in the Physics examination are normally distributed with a mean of 63 and a standard deviation of 10. Hee Jin also scored 80 marks in the Physics examination. [2 marks]
Find the probability that a randomly chosen candidate who sat the Physics examination scored **less** than Hee Jin.

15e. The candidates' marks in the Physics examination are normally distributed with a mean of 63 and a standard deviation of 10. Hee Jin also scored 80 marks in the Physics examination. [2 marks]
Determine whether Hee Jin's Physics mark, **compared to the other candidates**, is better than her mark in Chemistry. Give a reason for your answer.

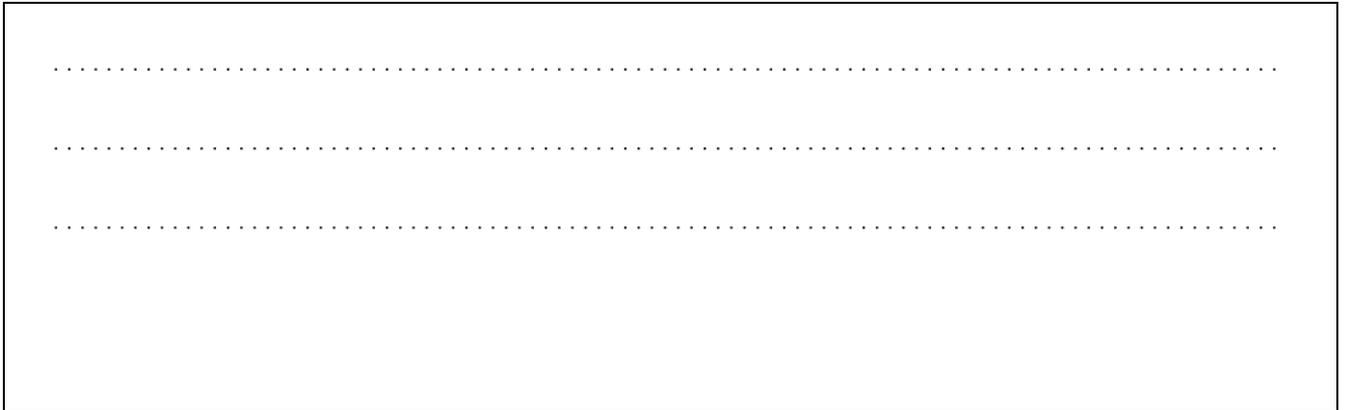
15f. To obtain a "grade A" a candidate must be in the top 10% of the candidates who sat the Physics examination. [3 marks]
Find the minimum possible mark to obtain a "grade A". Give your answer correct to the nearest integer.

A biologist is studying the relationship between the number of chirps of the Snowy Tree cricket and the air temperature. He records the chirp rate, x , of a cricket, and the corresponding air temperature, T , in degrees Celsius.

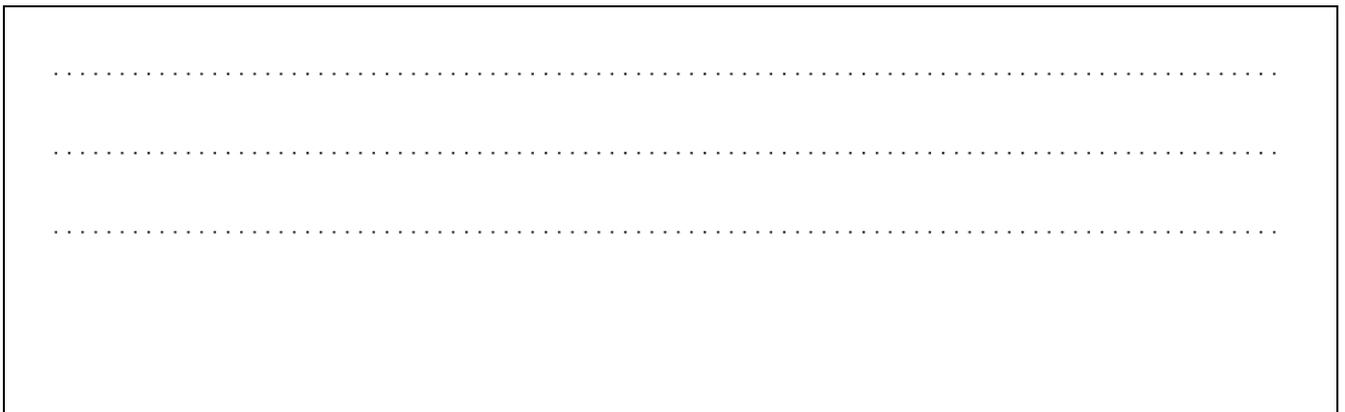
The following table gives the recorded values.

Cricket's chirp rate, x, (chirps per minute)	20	40	60	80	100	120
Temperature, T ($^{\circ}\text{C}$)	8.0	12.8	15.0	18.2	20.0	21.1

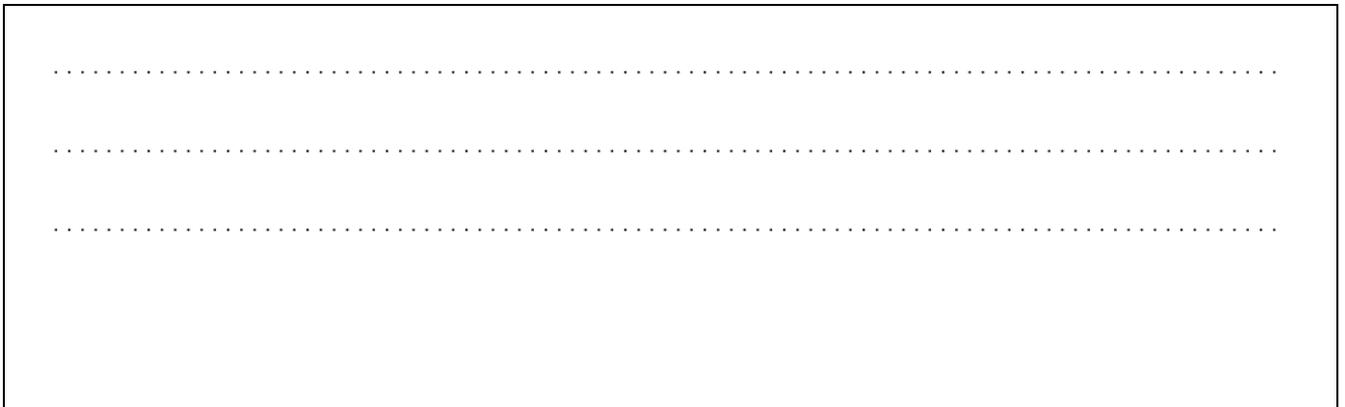
- 16a. Draw the scatter diagram for the above data. Use a scale of 2 cm for 20 chirps on the horizontal axis and 2 cm for 4°C on the vertical axis. [4 marks]



- 16b. Use your graphic display calculator to write down the Pearson's product-moment correlation coefficient, r , between x and T . [2 marks]



- 16c. Interpret the relationship between x and T using your value of r . [2 marks]



- 16d. Use your graphic display calculator to write down the equation of the regression line T on x . Give the equation [2 marks]
in the form $T = ax + b$.

.....

.....

.....

- 16e. Calculate the air temperature when the cricket's chirp rate is 70. [2 marks]

.....

.....

.....

- 16f. Given that $\bar{x} = 70$, draw the regression line T on x on your scatter diagram. [2 marks]

.....

.....

.....

- 16g. A forest ranger uses her own formula for estimating the air temperature. She counts the number of chirps in 15 seconds, z , multiplies this number by 0.45 and then she adds 10. [1 mark]

Write down the formula that the forest ranger uses for estimating the temperature, T .

Give the equation in the form $T = mz + n$.

.....

.....

.....

- 16h. A cricket makes 20 chirps in 15 seconds. [6 marks]

For this chirp rate

- (i) calculate an estimate for the temperature, T , **using the forest ranger's formula;**
- (ii) determine the actual temperature recorded by the biologist, **using the table above;**
- (iii) calculate the percentage error in the forest ranger's estimate for the temperature, compared to the actual temperature recorded by the biologist.

.....

.....

.....

A questionnaire was given to all members of a school community to find out which drink was the most popular to have with breakfast. The results are given in the table below, classified by age.

	Hot Chocolate	Tea	Coffee	Milk
Children aged 12 years and less	55	10	1	34
Teenagers aged from 13 to 19 years	25	35	20	10
Adults aged 20 years and over	20	40	79	6

A χ^2 test was conducted to decide whether the type of drink was independent of age.

- 17a. Find the number of degrees of freedom for the χ^2 test. [2 marks]

- 17b. Write down the null hypothesis for the χ^2 test. [1 mark]

- 17c. The critical value for the χ^2 test at the 5% significance level is 12.59. The χ^2 test statistic is calculated to be 146 with a p -value of 6.62×10^{-29} (both numbers given correct to 3 significant figures). [2 marks]

Write down the conclusion reached at the 5 % significance level. Give a clear reason for your answer.

A shop keeper recorded daily sales s of ice cream along with the daily maximum temperature t °C. The results for one week are shown below.

t	29	31	34	23	19	20	27
s	104	92	112	48	56	72	66

- 18a. Write down the equation of the regression line for s on t . [3 marks]

- 18b. Use your equation to predict the ice cream sales on a day when the maximum temperature is 24 °C. Give your answer correct to the nearest whole number. [3 marks]

The marks obtained by 8 candidates in Physics and Chemistry tests are given below.

Physics (x)	6	8	10	11	10	5	4	12
Chemistry (y)	8	11	14	13	11	7	5	15

- 19a. Write down the product moment correlation coefficient, r . [1 mark]

- 19b. Write down, in the form $y = mx + c$, the equation of the regression line y on x for the 8 candidates. [2 marks]

- 19c. A ninth candidate obtained a score of 7 in the Physics test but was absent for the Chemistry test. [2 marks]

Use your answer to (b) to estimate the score he would have obtained on the Chemistry test.

- 19d. Give a reason why it is valid to use this regression line to estimate the score on the Chemistry test in part (c). [1 mark]

Jorge conducted a survey of 200 drivers. He asked two questions:

- How long have you had your driving licence?
- Do you wear a seat belt when driving?

The replies are summarized in the table below.

	Wear a seat belt	Do not wear a seat belt
Licence less than 2 years	38	42
Licence between 2 and 15 years	30	45
Licence more than 15 years	30	15

20a. Jorge applies a χ^2 test at the 5% level to investigate whether wearing a seat belt is associated with the time a driver has had their licence. [8 marks]

- (i) Write down the null hypothesis, H_0 .
- (ii) Write down the number of degrees of freedom.
- (iii) Show that the expected number of drivers that wear a seat belt and have had their driving licence for more than 15 years is 22, correct to the nearest whole number.
- (iv) Write down the χ^2 test statistic for this data.
- (v) Does Jorge accept H_0 ? Give a reason for your answer.

20b. Consider the 200 drivers surveyed. One driver is chosen at random. Calculate the probability that [4 marks]

- (i) this driver wears a seat belt;
- (ii) the driver does not wear a seat belt, **given that** the driver has held a licence for more than 15 years.

20c. Two drivers are chosen at random. Calculate the probability that [6 marks]

- (i) both wear a seat belt.
- (ii) at least one wears a seat belt.

200 people of different ages were asked to choose their favourite type of music from the choices Popular, Country and Western and Heavy Metal. The results are shown in the table below.

Age/Music choice	Popular	Country and Western	Heavy Metal	Totals
11 – 25	35	5	50	90
26 – 40	30	10	20	60
41 – 60	20	25	5	50
Totals	85	40	75	200

It was decided to perform a chi-squared test for independence at the 5% level on the data.

21a. Write down the null hypothesis. [1 mark]

21b. Write down the number of degrees of freedom. [1 mark]

21c. Write down the chi-squared value. [2 marks]

21d. State whether or not you will reject the null hypothesis, giving a clear reason for your answer. [2 marks]

The number of bottles of water sold at a railway station on each day is given in the following table.

Day	0	1	2	3	4	5	6	7	8	9	10	11	12
Temperature (T°)	21	20.7	20	19	18	17.3	17	17.3	18	19	20	20.7	21
Number of bottles sold (n)	150	141	126	125	98	101	93	99	116	121	119	134	141

22a. Write down [2 marks]

- (i) the mean temperature;
- (ii) the standard deviation of the temperatures.

22b. Write down the correlation coefficient, [1 mark]

r , for the variables
 n and
 T .

22c. Comment on your value for [2 marks]

r .

22d. The equation of the line of regression for [2 marks]

n on
 T is
 $n = dT - 100$.

- (i) Write down the value of d .
- (ii) Estimate how many bottles of water will be sold when the temperature is 19.6° .

22e. On a day when the temperature was [1 mark]

36° Peter calculates that
314 bottles would be sold. Give one reason why his answer might be unreliable.

A survey of
400 people is carried out by a market research organization in two different cities, Buenos Aires and Montevideo. The people are asked which brand of cereal they prefer out of Chocos, Zucos or Fruti. The table below summarizes their responses.

	Chocos	Zucos	Fruti	Total
Buenos Aires	43	85	62	190
Montevideo	57	35	118	210
Total	100	120	180	400

23a. One person is chosen at random from those surveyed. Find the probability that this person [4 marks]

- (i) does not prefer Zucos;
- (ii) prefers Chocos, given that they live in Montevideo.

- 23b. Two people are chosen at random from those surveyed. Find the probability that they both prefer Fruti. [3 marks]
- 23c. The market research organization tests the survey data to determine whether the brand of cereal preferred is associated with a city. A chi-squared test at the 5% level of significance is performed. [1 mark]
State the null hypothesis.
- 23d. The market research organization tests the survey data to determine whether the brand of cereal preferred is associated with a city. A chi-squared test at the 5% level of significance is performed. [1 mark]
State the number of degrees of freedom.
- 23e. The market research organization tests the survey data to determine whether the brand of cereal preferred is associated with a city. A chi-squared test at the 5% level of significance is performed. [2 marks]
Show that the expected frequency for the number of people who live in Montevideo and prefer Zucos is 63.
- 23f. The market research organization tests the survey data to determine whether the brand of cereal preferred is associated with a city. A chi-squared test at the 5% level of significance is performed. [2 marks]
Write down the chi-squared statistic for this data.
- 23g. The market research organization tests the survey data to determine whether the brand of cereal preferred is associated with a city. A chi-squared test at the 5% level of significance is performed. [2 marks]
State whether the market research organization would accept the null hypothesis. Clearly justify your answer.

The following table shows the cost in AUD of seven paperback books chosen at random, together with the number of pages in each book.

Book	1	2	3	4	5	6	7
Number of pages (x)	50	120	200	330	400	450	630
Cost (y AUD)	6.00	5.40	7.20	4.60	7.60	5.80	5.20

- 23h. Plot these pairs of values on a scatter diagram. Use a scale of 1 cm to represent 50 pages on the horizontal axis and 1 cm to represent 1 AUD on the vertical axis. [3 marks]
- 23i. Write down the linear correlation coefficient, r , for the data. [2 marks]
- 23j. Stephen wishes to buy a paperback book which has 350 pages in it. He plans to draw a line of best fit to determine the price. State whether or not this is an appropriate method in this case and justify your answer. [2 marks]

