Topic 4 Part 2 *[189 marks]*

Part A

100 students are asked what they had for breakfast on a particular morning. There were three choices: cereal (X) , bread (Y) and fruit (Z). It is found that

10 students had all three

17 students had bread and fruit only

15 students had cereal and fruit only

12 students had cereal and bread only

13 students had only bread

8 students had only cereal

9 students had only fruit

1a. Represent this information on a Venn diagram. *[4 marks]*

1b. Find the number of students who had none of the three choices for breakfast. *[2 marks]*

1c. Write down the percentage of students who had fruit for breakfast. *[2 marks]*

1d. Describe in words what the students in the set $X \cap Y'$ had for breakfast. *[2 marks]*

1e. Find the probability that a student had **at least** two of the three choices for breakfast. *[2 marks]*

1f. Two students are chosen at random. Find the probability that both students had all three choices for breakfast. *[3 marks]*

Part B

The same 100 students are also asked how many meals on average they have per day. The data collected is organized in the following table.

	3 or fewer meals per day	4 or 5 meals per day	More than 5 meals per day	Total
Male	15	25	15	55
Female	12	20	13	45
Total	27	45	28	100

A
χ^2 test is carried out at the 5 % level of significance.

1g. Write down the null hypothesis, H_0, for this test. *[1 mark]*

1h. Write down the number of degrees of freedom for this test. *[1 mark]*

1i. Write down the critical value for this test. *[1 mark]*

1j. Show that the expected number of females that have more than 5 meals per day is 13, correct to the nearest integer. *[2 marks]*

1k. Use your graphic display calculator to find the χ^2_{calc} for this data. *[2 marks]*

1l. Decide whether H_0 must be accepted. Justify your answer. *[2 marks]*

Pam has collected data from a group of 400 IB Diploma students about the Mathematics course they studied and the language in which they were examined (English, Spanish or French). The summary of her data is given below.

	Mathematics HL	Mathematics SL	Mathematical Studies SL	**Total**
English	50	70	80	200
Spanish	30	50	30	110
French	20	30	40	90
Total	100	150	150	400

2a. A student is chosen at random from the group. Find the probability that the student *[8 marks]*

(i) studied Mathematics HL;

(ii) was examined in French;

(iii) studied Mathematics HL and was examined in French;

(iv) did not study Mathematics SL and was not examined in English;

(v) studied Mathematical Studies SL given that the student was examined in Spanish.

2b. Pam believes that the Mathematics course a student chooses is independent of the language in which the student is examined. *[2 marks]*

Using your answers to parts (a) (i), (ii) and (iii) above, state whether there is any evidence for Pam's belief. Give a reason for your answer.

2c. Pam decides to test her belief using a Chi-squared test at the 5% level of significance. *[3 marks]*

(i) State the null hypothesis for this test.

(ii) Show that the expected number of Mathematical Studies SL students who took the examination in Spanish is 41.3, correct to 3 significant figures.

2d. Write down *[4 marks]*

(i) the Chi-squared calculated value;

(ii) the number of degrees of freedom;

(iii) the Chi-squared critical value.

2e. State, giving a reason, whether there is sufficient evidence at the 5% level of significance that Pam's belief is correct. *[2 marks]*

The heat output in thermal units from burning
1 kg of wood changes according to the wood's percentage moisture content. The moisture content and heat output of
10 blocks of the same type of wood each weighing
1 kg were measured. These are shown in the table.

Moisture content % (x)	8	15	22	30	34	45	50	60	74	82
Heat output (y)	80	77	74	69	68	61	61	55	50	45

3a. Draw a scatter diagram to show the above data. Use a scale of *[4 marks]*
2 cm to represent
10% on the x-axis and a scale of
2 cm to represent
10 thermal units on the y-axis.

3b. Write down *[2 marks]*
(i) the mean percentage moisture content,
$\bar{x}$;
(ii) the mean heat output,
$\bar{y}$.

3c. Plot the point *[2 marks]*
$(\bar{x}, \bar{y})$ on your scatter diagram and label this point M .

3d. Write down the product-moment correlation coefficient, *[2 marks]*
r .

3e. The equation of the regression line *[2 marks]*
y on
x is
$y = -0.470x + 83.7$. Draw the regression line
y on
x on your scatter diagram.

3f. The equation of the regression line *[2 marks]*
y on
x is
$y = -0.470x + 83.7$. Estimate the heat output in thermal units of a
1 kg block of wood that has
25% moisture content.

3g. The equation of the regression line *[2 marks]*
y on
x is
$y = -0.470x + 83.7$. State, with a reason, whether it is appropriate to use the regression line
y on
x to estimate the heat output in part (f).

One day the numbers of customers at three cafés, “Alan’s Diner” (A), “Sarah’s Snackbar” (S) and “Pete’s Eats” (P), were recorded and are given below.

17 were customers of Pete’s Eats only
27 were customers of Sarah’s Snackbar only
15 were customers of Alan’s Diner only
10 were customers of Pete’s Eats **and** Sarah’s Snackbar **but not** Alan’s Diner
8 were customers of Pete’s Eats **and** Alan’s Diner **but not** Sarah’s Snackbar

4a. Draw a Venn Diagram, using sets labelled A, S and P, that shows this information. *[3 marks]*

4b. There were 48 customers of Pete’s Eats that day. Calculate the number of people who were customers of all three cafés. *[2 marks]*

4c. There were 50 customers of Sarah’s Snackbar that day. Calculate the total number of people who were customers of Alan’s Diner. *[3 marks]*

4d. Write down the number of customers of Alan’s Diner that were also customers of Pete’s Eats. *[1 mark]*

4e. Find $n[(S \cup P) \cap A']$. *[2 marks]*

Some of the customers in each café were given survey forms to complete to find out if they were satisfied with the standard of service they received.

	Pete’s Eats	Alan’s Diner	Sarah’s Snackbar	Total
Dissatisfied	16	8	16	40
Satisfied	26	20	34	80
Total	42	28	50	120

4f. One of the survey forms was chosen at random, find the probability that the form showed “Dissatisfied”; *[2 marks]*

4g. One of the survey forms was chosen at random, find the probability that the form showed “Satisfied” and was completed at Sarah’s Snackbar; *[2 marks]*

4h. One of the survey forms was chosen at random, find the probability that the form showed “Dissatisfied”, given that it was completed at Alan’s Diner. *[2 marks]*

4i. A χ^2 test at the 5% significance level was carried out to determine whether there was any difference in the level of customer satisfaction in each of the cafés. *[1 mark]*

Write down the null hypothesis, H_0, for the χ^2 test.

4j. A χ^2 test at the 5% significance level was carried out to determine whether there was any difference in the level of customer satisfaction in each of the cafés. *[1 mark]*

Write down the number of degrees of freedom for the test.

4k. A χ^2 test at the 5% significance level was carried out to determine whether there was any difference in the level of customer satisfaction in each of the cafés. *[2 marks]*

Using your graphic display calculator, find χ^2_{calc}.

4l. A χ^2 test at the 5% significance level was carried out to determine whether there was any difference in the level of customer satisfaction in each of the cafés. *[2 marks]*

State, giving a reason, the conclusion to the test.

Part A

A university required all Science students to study one language for one year. A survey was carried out at the university amongst the 150 Science students. These students all studied one of either French, Spanish or Russian. The results of the survey are shown below.

	French	**Spanish**	**Russian**
Female	9	29	12
Male	31	40	29

Ludmila decides to use the χ^2 test at the 5% level of significance to determine whether the choice of language is independent of gender.

5a. State Ludmila's null hypothesis. *[1 mark]*

5b. Write down the number of degrees of freedom. *[1 mark]*

5c. Find the expected frequency for the females studying Spanish. *[2 marks]*

5d. Use your graphic display calculator to find the χ^2 test statistic for this data. *[2 marks]*

5e. State whether Ludmila accepts the null hypothesis. Give a reason for your answer. *[2 marks]*

At the end of the year, only seven of the female Science students sat examinations in Science and French. The marks for these seven students are shown in the following table.

Science (S)	23	51	56	62	12	73	72
French (F)	65	45	45	40	70	36	30

5f. Draw a labelled scatter diagram for this data. Use a scale of 2 cm to represent 10 marks on the x-axis (S) and 10 marks on the y-axis (F). *[4 marks]*

5g. Use your graphic calculator to find *[2 marks]*

(i) $\bar{S}$, the mean of S;

(ii) $\bar{F}$, the mean of F.

5h. Plot the point $\mathrm{M}(\bar{S}, \bar{F})$ on your scatter diagram. *[1 mark]*

5i. Use your graphic display calculator to find the equation of the regression line of F on S. *[2 marks]*

5j. Draw the regression line on your scatter diagram. *[2 marks]*

5k. Carletta's mark on the Science examination was 44. She did not sit the French examination. *[2 marks]*

Estimate Carletta's mark for the French examination.

5l. Monique's mark on the Science examination was 85. She did not sit the French examination. Her French teacher wants to use the regression line to estimate Monique's mark. *[2 marks]*

State whether the mark obtained from the regression line for Monique's French examination is reliable. Justify your answer.

Tony wants to carry out a χ^2 test to determine whether or not a person's choice of one of the three professions; engineering, medicine or law is influenced by the person's sex (gender).

6a. State the null hypothesis, H_0, for this test. *[1 mark]*

6b. Write down the number of degrees of freedom. *[1 mark]*

6c. Of the 400 people Tony interviewed, 220 were male and 180 were female. 80 of the people had chosen engineering as a profession. *[2 marks]*

Calculate the expected number of female engineers.

6d. Tony used a 5 % level of significance for his test and obtained a p-value of 0.0634 correct to 3 significant figures. *[2 marks]*

State Tony's conclusion to the test. Give a reason for this conclusion.

A market researcher consulted males and females to determine whether the type of coffee they drink is associated with gender. The types of coffee are Cappuccino, Latte, Americano, Macchiato and Espresso. A χ^2 test was conducted, at the 5 % significance level and the χ^2 value was found to be 8.73.

7a. Write down the null hypothesis. *[1 mark]*

7b. Write down the alternative hypothesis. *[1 mark]*

7c. Write down the number of degrees of freedom for this test. *[1 mark]*

7d. Write down the critical value for this test. *[1 mark]*

7e. State whether the type of coffee drunk is independent of gender. Give a reason for your answer. *[2 marks]*

In an environmental study of plant diversity around a lake, a biologist collected data about the number of different plant species (y) that were growing at different distances (x) in metres from the lake shore.

Distance (x)	2	5	8	10	13	17	23	35	40
Plant species (y)	35	34	30	29	24	19	15	13	8

8a. Draw a scatter diagram to show the data. Use a scale of 2 cm to represent 10 metres on the x-axis and 2 cm to represent 10 plant species on the y-axis. *[4 marks]*

8b. Using your scatter diagram, describe the correlation between the number of different plant species and the distance from the lake shore. *[1 mark]*

8c. Use your graphic display calculator to write down $\bar{x}$, the mean of the distances from the lake shore. *[1 mark]*

8d. Use your graphic display calculator to write down $\bar{y}$, the mean number of plant species. *[1 mark]*

8e. Plot the point ($\bar{x}$, $\bar{y}$) on your scatter diagram. **Label this point M.** *[2 marks]*

8f. Write down the equation of the regression line y on x for the above data. *[2 marks]*

8g. Draw the regression line y on x on your scatter diagram. *[2 marks]*

8h. Estimate the number of plant species growing 50 metres from the lake shore. *[2 marks]*

Alex and Kris are riding their bicycles together along a bicycle trail and note the following distance markers at the given times.

Time (t hours)	1	2	3	4	5	6	7
Distance (d km)	57	65	72	81	89	97	107

9a. Draw a scatter diagram of the data. Use 1 cm to represent 1 hour and 1 cm to represent 10 km. *[3 marks]*

9b. Write down for this set of data the mean time, $\bar{t}$. *[1 mark]*

9c. Write down for this set of data the mean distance, $\bar{d}$. *[1 mark]*

9d. Mark and label the point $M(\bar{t}, \bar{d})$ on your scatter diagram. *[2 marks]*

9e. Draw the line of best fit on your scatter diagram. *[2 marks]*

9f. **Using your graph**, estimate the time when Alex and Kris pass the 85 km distance marker. Give your answer correct to **one decimal place**. *[2 marks]*

9g. Write down the equation of the regression line for the data given. *[2 marks]*

9h. **Using your equation** calculate the distance marker passed by the cyclists at 10.3 hours. *[2 marks]*

9i. Is this estimate of the distance reliable? Give a reason for your answer. *[2 marks]*

In a mountain region there appears to be a relationship between the number of trees growing in the region and the depth of snow in winter. A set of 10 areas was chosen, and in each area the number of trees was counted and the depth of snow measured. The results are given in the table below.

Number of trees (x)	**Depth of snow in cm (y)**
45	30
75	50
66	40
27	25
44	30
28	5
60	35
35	20
73	45
47	25

10a. Use your graphic display calculator to find the mean number of trees. *[1 mark]*

10b. Use your graphic display calculator to find the mean depth of snow. *[1 mark]*

10c. Use your graphic display calculator to find the standard deviation of the depth of snow. *[1 mark]*

10d. The covariance, $S_{xy} = 188.5$. *[2 marks]*

Write down the product-moment correlation coefficient, r.

10e. Write down the equation of the regression line of y on x. *[2 marks]*

10f. If the number of trees in an area is 55, estimate the depth of snow. *[2 marks]*

10g. Use the equation of the regression line to estimate the depth of snow in an area with 100 trees. *[1 mark]*

10h. Decide whether the answer in (e)(i) is a valid estimate of the depth of snow in the area. Give a reason for your answer. *[2 marks]*

In a study on
100 students there seemed to be a difference between males and females in their choice of favourite car colour. The results are given in the table below. A
χ^2 test was conducted.

	Blue	Red	Green
Males	14	6	8
Females	31	24	17

10i. Write down the total number of male students. *[1 mark]*

10j. Show that the expected frequency for males, whose favourite car colour is blue, is 12.6. *[2 marks]*

10k. The calculated value of *[1 mark]*
χ^2 is
1.367 and the critical value of
χ^2 is
5.99 at the
5% significance level.

Write down the null hypothesis for this test.

10l. The calculated value of *[1 mark]*
χ^2 is
1.367 and the critical value of
χ^2 is
5.99 at the
5% significance level.

Write down the number of degrees of freedom.

10m. The calculated value of *[2 marks]*
χ^2 is
1.367 and the critical value of
χ^2 is
5.99 at the
5% significance level.

Determine whether the null hypothesis should be accepted at the
5% significance level. Give a reason for your answer.

A manufacturer claims that fertilizer has an effect on the height of rice plants. He measures the height of fertilized and unfertilized plants. The results are given in the following table.

Plant height	Fertilized plants	Unfertilized plants
>75 cm	115	80
50 – 75 cm	45	65
<50 cm	20	35

A chi-squared test is performed to decide if the manufacturer's claim is justified at the **1 %** level of significance.

11a. Write down the null and alternative hypotheses for this test. *[2 marks]*

11b. For the number of fertilized plants with height greater than 75 cm, show that the expected value is 97.5. *[3 marks]*

11c. Write down the value of χ^2_{calc}. *[2 marks]*

11d. Write down the number of degrees of freedom. *[1 mark]*

11e. Is the manufacturer's claim justified? Give a reason for your answer. *[2 marks]*

The population of fleas on a dog after t days, is modelled by

$$N = 4 \times (2)^{\frac{t}{4}}, \ t \geqslant 0$$

Some values of N are shown in the table below.

t	0	4	8	12	16	20
N	p	8	16	32	q	128

11f. Write down the value of p. *[1 mark]*

11g. Write down the value of q. *[2 marks]*

11h. Using the values in the table above, draw the graph of N for $0 \leq t \leq 20$. Use 1 cm to represent 2 days on the horizontal axis and 1 cm to represent 10 fleas on the vertical axis. *[6 marks]*

11i. **Use your graph** to estimate the number of days for the population of fleas to reach 55. *[2 marks]*