9 Samantha wanted to find out whether there was a connection between the type of degree that a person had and their annual salary in dollars. She interviewed 120 professionals and her observed results are shown in the following table.

Degree	BA	MA	PhD	Totals
< US $60 000	17	8	4	29
US $60 000–US $120 000	14	19	9	42
> US $120 000	8	14	27	49
Totals	39	41	40	120

Test, at the 5% significance level, whether there is a connection between degree and salary.

a State the null hypothesis and the alternative hypothesis.

b Write down the number of degrees of freedom.

c Write down the χ^2 test statistic and the p-value for this data.

The critical value is 9.488.

d Comment on your result.

Developing inquiry skills

Let the measurement of the largest height of the trees in the opening problem be h. Divide the trees into small, medium and large, where small trees have $h \leq 4.5$ m, medium, $4.5 < h \leq 5.0$ m, and large $h > 5.0$ m.

Use these categories to form a contingency table for the trees in areas A and B, and test at the 5% significance level whether the heights of the trees are independent of the forest area they were taken from.

Does the conclusion of the test support the hypothesis that the trees from area A, on average, are taller than those from area B? Justify your answer.

8.3 χ^2 goodness of fit test

Investigation 4

Jiang wonders whether the die he was given is fair. He rolls it 300 times. His results are shown in the table.

Number	Frequency
1	35
2	52
3	47
4	71
5	62
6	33

TOK

How have technological innovations affected the nature and practice of mathematics?

1 Write down the probability of throwing a 1 on a fair die.

2 If you throw a fair die 300 times, how many times would you expect to throw a 1?

3 Write down the expected frequencies for throwing a fair die 300 times.

Number	Expected frequency
1	
2	
3	
4	
5	
6	

4 **Factual** Do you need to combine any rows on your table?

Since all the expected frequencies are the same, this is known as a **uniform distribution.**

5 **Factual** Is the formula for the χ^2 test suitable to test whether Jiang's results fit this uniform distribution?

The null hypothesis is H_0 : Jiang's die satisfies a uniform distribution.

6 Write down the alternative hypothesis.

7 Given that the critical value at the 5% significance level for this test is 11.07, use the formula for the χ^2 test, $\chi^2_{\text{calc}} = \sum \frac{(f_o - f_e)^2}{f_e}$, to find out whether Jiang's results could be taken from a uniform distribution.

Normally you would solve this using your GDC, which may ask you to enter the degrees of freedom.

8 **Factual** What is the number of degrees of freedom in a χ^2 goodness of fit test? (Consider in how many cells you have free choices when completing the expected values table.)

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

10 Using your GDC, find the test statistic and the p-value.

11 What is your conclusion from this test?

12 **Conceptual** What is the purpose of the χ^2 goodness of fit test?

These types of test are called "goodness of fit" tests as they are measuring how closely the observed data fits with the expected data for a particular distribution. The test for independence using contingency tables is an example of a goodness of fit test, but you can test for the goodness of fit for any distribution.

In a χ^2 goodness of fit test, the number of degrees of freedom is $\nu = (n - 1)$.

Example 4

The students in Year 8 are asked what day of the week their birthdays are on this year. The table shows the results.

Day	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Frequency	12	14	18	17	15	15	14

a Write down the table of expected values, given that each day is equally likely.

b Conduct a χ^2 goodness of fit test at the 5% significance level for this data.

c The critical value is 12.592. Write down the conclusion for the test.

a

Day	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Frequency	15	15	15	15	15	15	15

b H_0 : The data satisfies a uniform distribution.

H_1 : The data does not satisfy a uniform distribution.

$\nu = (7 - 1) = 6$

Using GDC, $\chi^2 = 1.60$ and p-value = 0.953.

c $0.95 > 0.05$ or $1.60 < 12.592$, so you can accept the null hypothesis: the data does satisfy a uniform distribution.

Exercise 8D

1 Terri buys 10 packets of Skittles and counts how many of each colour (yellow, orange, red, purple and green) there are. In total she has 600 sweets.

According to the Skittles website, the colours should be evenly distributed with 20% of each colour in a bag.

The results for Terri's 10 bags are:

Colour	Frequency
Yellow	104
Orange	132
Red	98
Purple	129
Green	137

a Find the expected frequencies.

b Write down the number of degrees of freedom.

c Determine the results of a goodness of fit test at the 5% significance level to find out whether Terri's data fits a uniform distribution. Remember to write down the null and alternative hypotheses.

The critical value for this test is 9.488.

d State the conclusion for the test and give a reason for your answer.

2 There are 60 students in Grade 12. Mr Stewart asks them which month their birthdays are in, and the results are shown in the table.

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Frequency	3	5	4	6	5	6	4	7	8	6	3	3

The months in which people have birthdays are uniformly distributed.

a Write down the table of expected values.

b Write down the number of degrees of freedom.

c Determine the results of a goodness of fit test at the 10% significance level to find out whether the data fits a uniform distribution. Remember to write down the null and alternative hypotheses.

The critical value for this test is 17.275.

d State the conclusion for the test and give a reason for your answer.

3 Sergei works in a call centre. One week he answers 840 calls. The number of calls that he answers each day are shown in the table.

Day	Mon	Tues	Wed	Thurs	Fri	Sat	Sun
Frequency	146	98	103	106	93	204	90

The calls are uniformly distributed.

a Show that the expected value of the number of calls each day is 120.

b Write down the number of degrees of freedom.

c Determine the results of a goodness of fit test at the 5% significance level to find out whether the data fits a uniform distribution. Remember to write down the null and alternative hypotheses.

The critical value for this test is 12.592.

d State the conclusion for the test and give a reason for your answer.

4 The last digit on 500 winning lottery tickets is recorded in the table.

Last digit	0	1	2	3	4	5	6	7	8	9
Frequency	44	53	49	61	47	52	39	58	42	45

a Each number should be equally likely to occur. Write down the table of expected values.

b Write down the number of degrees of freedom.

c Determine the results of a goodness of fit test at the 10% significance level to find out whether the data fits a uniform distribution. Remember to write down the null and alternative hypotheses.

The critical value for this test is 14.684.

d State the conclusion for the test and give a reason for your answer.

Example 5

The scores for IQ tests are normally distributed with a mean of 100 and standard deviation of 10. Cinzia gives an IQ test to all 200 IB Diploma Programme students in the school. Her results are shown in the table.

Score, x	Frequency
$x < 90$	5
$90 \leq x < 100$	14
$100 \leq x < 110$	74
$110 \leq x < 120$	58
$120 \leq x < 130$	34
$130 \leq x$	15

Cinzia wants to test whether these results are also normally distributed and performs a χ^2 goodness of fit test at the 10% significance level.

a Write down her null and alternative hypotheses.

b Find the expected values.

c If any expected values are less than 5 then rewrite both tables.

d Write down the number of degrees of freedom.

The critical value is 6.251.

e Find the χ^2 test statistic and the p-value, and state the conclusion for the test.

a H_0 : The scores are normally distributed with mean of 100 and standard deviation of 10.

H_1 : The scores are not normally distributed with mean of 100 and standard deviation of 10.

Continued on next page

b

Score, x	Probability	Expected score
$x<90$	0.1587	31.7
$90\leq x<100$	0.3413	68.3
$100\leq x<110$	0.3413	68.3
$110\leq x<120$	0.1359	27.2
$120\leq x<130$	0.0214	4.28
$130\leq x$	0.00135	0.270

Use the normal cdf function on your GDC to find the probability and then multiply your answer by 200.

c The last two scores are both less than 5, even when added together, and so you will have to combine them with the one above.

Observed and expected values are now:

Score, x	Observed score	Expected score
$x<90$	5	31.7
$90\leq x<100$	14	68.3
$100\leq x<110$	74	68.3
$110\leq x$	107	31.7

$4.28 + 0.270$ is still less than 5.

d $v = 3$

$v = (4 - 1)$

e χ^2 value is 245 and the p-value is 7.89×10^{-53}. The p-value is less than 0.10 and so the null hypothesis is rejected: the scores are not normally distributed with mean of 100 and standard deviation of 10.

Also, the critical value for $v = 3$ at the 10% significance level is 6.251. $245 > 6.251$, so the null hypothesis is rejected.

Exercise 8E

1 Marius works in a fish shop. One week he measures 250 fish before selling them. His results are shown in the table.

Length of fish, x cm	Frequency
$9\leq x<12$	5
$12\leq x<15$	22
$15\leq x<18$	71
$18\leq x<21$	88
$21\leq x<24$	52
$24\leq x<27$	10
$27\leq x<30$	2

Marius is told that the lengths of the fish should be normally distributed with a mean of 19 cm and standard deviation of 3 cm, so he decides to perform a χ^2 goodness of fit test at the 5% significance level to find out whether the fish that he measured could have come from a population with this distribution.

a Write down his null and alternative hypotheses.

b Find the probability that a fish is between 9 cm and 12 cm.

c In total, 250 fish were measured. Calculate how many fish you expect to be between 9 cm and 12 cm.

d Complete the table of expected values for 250 normally distributed fish with a mean of 19 cm and standard deviation of 3 cm.

Length of fish, x cm	Probability	Expected frequency
$9 \le x < 12$	0.009 386	2.35
$12 \le x < 15$	0.0814	20.3
$15 \le x < 18$		
$18 \le x < 21$		
$21 \le x < 24$		
$24 \le x < 27$	0.043 96	10.99
$27 \le x < 30$	0.003 71	0.927

Two of the expected frequencies are less than 5.

e Discuss what you have to do in this case. Remember that this may be encountered in internal assessments, but will not be in examinations.

f Rewrite the original table and the table of expected values so that all the expected values are greater than 5.

g Write down the number of degrees of freedom.

h Find the χ^2 value and the p-value.

The critical value is 9.488.

i Write down your conclusion for this test.

2 The weights of sample group are normally distributed with a mean of 52 kg and standard deviation of 3 kg.

The district nurse weighs 200 people and her results are shown in the table.

Weight, w kg	$w < 45$	$45 \le w < 50$	$50 \le w < 55$	$55 \le w < 60$	$w \ge 60$
Frequency	12	44	82	53	9

a Complete the expected frequency table.

Weight, w kg	$w < 45$	$45 \le w < 50$	$50 \le w < 55$	$55 \le w < 60$	$w \ge 60$
Expected frequency	1.96	48.54			0.77

b Rewrite the table of observed frequencies so that all the expected frequencies are greater than 5 and find the corresponding expected frequencies.

c Write down the number of degrees of freedom.

d Determine the results of a goodness of fit test at the 5% significance level to find out whether the data fits a normal distribution. Remember to write down the null and alternative hypotheses.

The critical value for this test is 5.991.

e State the conclusion for the test and give a reason for your answer.

3 The grades for an economics exam for 300 university students are as follows.

Grade, x%	$x < 50$	$50 \le x < 60$	$60 \le x < 70$	$70 \le x < 80$	$x \ge 80$
Frequency	8	72	143	71	6

The grades are normally distributed with a mean of 65% and standard deviation of 7.5%.

a Complete the expected frequency table.

Grade, x%	$x < 50$	$50 \le x < 60$	$60 \le x < 70$	$70 \le x < 80$	$x \ge 80$
Expected frequency	6.8			68.92	6.8

b Write down the number of degrees of freedom.

c Determine the results of a goodness of fit test at the 10% significance level to find out whether the data fits a normal distribution. Remember to write down the null and alternative hypotheses.

The critical value for this test is 7.779.

d State the conclusion for the test and give a reason for your answer.

4 The heights of elephants are normally distributed with a mean of 250 cm and standard deviation of 11 cm. Two hundred and fifty elephants are measured; the results are shown in the table.

Height, h cm	$h < 235$	$235 \le h < 245$	$245 \le h < 255$	$255 \le h < 265$	$h \ge 265$
Frequency	10	69	88	63	20

a Complete the expected frequency table.

Height, h cm	$h < 235$	$235 \le h < 245$	$245 \le h < 255$	$255 \le h < 265$	$h \ge 265$
Expected frequency	21.6				

b Write down the number of degrees of freedom.

c Determine the results of a goodness of fit test at the 5% significance level to find out whether the data fits a normal distribution. Remember to write down the null and alternative hypotheses.

The critical value for this test is 9.488.

d State the conclusion for the test and give a reason for your answer.

5 The lifespan of light bulbs is normally distributed with a mean lifespan of 1200 hours and standard deviation of 100 hours. Four hundred light bulbs are tested and the results are shown in the table.

Lifespan, h hours	$h<1000$	$1000\leq h<1100$	$1100\leq h<1200$	$1200\leq h<1300$	$1300\leq h<1400$	$h\geq 1400$
Frequency	24	52	92	164	42	26

a Complete the expected frequency table.

Lifespan, h hours	$h<1000$	$1000\leq h<1100$	$1100\leq h<1200$	$1200\leq h<1300$	$1300\leq h<1400$	$h\geq 1400$
Expected frequency	9.1					

b Write down the number of degrees of freedom.

c Determine the results of a goodness of fit test at the 5% significance level to find out whether the data fits a normal distribution. Remember to write down the null and alternative hypotheses.

The critical value for this test is 11.070.

d State the conclusion for the test and give a reason for your answer.

Example 6

Using what you learned in Chapter 7, find the probability when you toss three coins of obtaining: 0 heads, exactly 1 head, exactly 2 heads, 3 heads.

Hagar tosses three coins 200 times and makes a note of the number of heads each time. Her results are as follows.

Number of heads	Frequency
0	28
1	67
2	83
3	22

She is interested in finding out whether her results follow a binomial distribution and performs a χ^2 goodness of fit test at the 5% significance level.

a Using the terms of B(3, 0.5) and the fact that Hagar tossed the coins 200 times, find the expected values for the number of heads.

b Comment on whether any of these values are less than 5.

c Write down the null and alternative hypotheses and the degrees of freedom.

The critical value is 7.815.

d Find the χ^2 value and the p-value.

e Write down the conclusion for this test.

TOK

To what extent can shared knowledge be distorted and misleading?

a

Number of heads	Expected frequency
0	$200 \times 0.125 = 25$
1	$200 \times 0.375 = 75$
2	$200 \times 0.375 = 75$
3	$200 \times 0.125 = 25$

b They are all greater than 5.

c H_0 : The number of heads follows a binomial distribution.

H_1 : The number of heads does not follow a binomial distribution.

d The χ^2 value is 2.426... and the p-value is 0.4886...

e $2.426... < 7.815$, so the null hypothesis is accepted. Or, $0.4886... > 0.05$, so the null hypothesis is accepted. The number of heads follows a binomial distribution.

International-mindedness

The physicist Frank Oppenheimer wrote:

"Prediction is dependent only on the assumption that observed patterns will be repeated."

This is the danger of extrapolation. There are many examples of its failure in the past: for example share prices, the spread of disease and climate change.

Exercise 8F

1 Percy sews three seeds in each of 50 different pots. The probability that a seed will germinate is 0.75. The number of seeds that germinated in each pot is shown in the table.

Number of seeds germinating	0	1	2	3
Frequency	5	10	15	20

a Using the binomial expansion B(3, 0.75), find the expected probabilities of 0, 1, 2 or 3 seeds germinating.

b Write down the table of expected frequencies.

c State whether or not there are any expected values less than 5.

d Write down the number of degrees of freedom.

e Determine the results of a goodness of fit test at the 5% significance level to find out whether the data fits a binomial distribution. Remember to write down the null and alternative hypotheses.

The critical value for this test is 5.991.

f State the conclusion for the test and give a reason for your answer.

2 The number of boys in 100 families with three children is shown in the table.

Number of boys	0	1	2	3
Frequency	16	23	32	19

a If the probability of having a boy is 0.5, use the binomial expansion B(3, 0.5) to find the expected values.

b State whether or not there are any expected values less than 5.

c Write down the number of degrees of freedom.

d Determine the results of a goodness of fit test at the 1% significance level to find out whether the data fits a binomial distribution. Remember to write down the null and alternative hypotheses.

The critical value for this test is 11.345.

e State the conclusion for the test and give a reason for your answer.

3 Esmerelda tosses two unbiased dice 250 times. She records the number of 6s that she tosses.

Number of 6s	0	1	2
Frequency	135	105	10

a Use the binomial expansion $B(2, \frac{1}{6})$ to complete the table of expected values.

Number of 6s	0	1	2
Expected frequencies			6.94

b State whether there are any expected values less than 5.

c Write down the degrees of freedom.

d Determine the results of a goodness of fit test at the 5% significance level to find out whether the data fits a binomial distribution. Remember to write down the null and alternative hypotheses.

The critical value for this test is 5.991.

e State the conclusion for the test and give a reason for your answer.

4 A multiple-choice test has five questions. Each question has four answers to choose from.

a Find the probability of getting any one question correct.

Five hundred students sit the test and the results are shown in the table.

Number correct, n	0	1	2	3	4	5
Frequency	38	66	177	132	51	36

b Using the binomial expansion B(5, 0.25), find the expected probabilities of having 0, 1, 2, 3, 4 or 5 questions correct.

c Complete the table of expected frequencies.

Number correct, n	0	1	2	3	4	5
Expected frequency					7.32	

d State whether there are any expected values less than 5.

e Write down the number of degrees of freedom.

f Determine the results of a goodness of fit test at the 5% significance level to find out whether the data fits a binomial distribution. Remember to write down the null and alternative hypotheses.

The critical value for this test is 9.488.

g State the conclusion for the test and give a reason for your answer.

Reflect What does the χ^2 goodness of fit test do?

Developing inquiry skills

	Area A	Area B
Small ($h \leq 4.5$)	3	9
Medium ($4.5 < h \leq 5.0$)	7	9
Large ($h > 5.0$)	14	6

From previous research it is known that this species of tree as a whole follows a normal distribution with a mean of 4.9 m and a standard deviation of 0.5 m.

Test the trees from each of the forest areas separately and see whether the observed values are consistent with both samples being taken from this distribution.

Combine the trees from both areas and carry out the test again.

What do your results suggest about the likelihood of the trees from area A being taller than those from area B?