Question 1

At a school in Copenhagen, it is believed that favourite music genre is related to gender. 400 students were asked to indicate their favourite music genre from a selection and the results are indicated in the table below.

	Pop	Rock	Classical	Rap
Female	58	63	17	44
Male	23	96	12	87

It is decided to test this hypothesis by using a χ^2 test at the 5% significance level.
The critical value is 7.815.

(a) State the null and alternative hypotheses for this test.

[2]

(b) Write down the number of degrees of freedom for this table.

[1]

(c) Calculate the χ^2 test statistic for this data.

[2]

(d) What conclusion can be drawn from this test? Give a reason for your answer.

[2]

a) H_0 : The favourite music genre is independent of gender.

H_1 : The favourite music genre is not independent of gender.

At a school in Copenhagen, it is believed that favourite music genre is related to gender.
400 students were asked to indicate their favourite music genre from a selection and the results are indicated in the table below.

4

	Pop	Rock	Classical	Rap
Female	58	63	17	44
Male	23	96	12	87

2

It is decided to test this hypothesis by using a χ^2 test at the 5% significance level.
The critical value is 7.815.

(a) State the null and alternative hypotheses for this test.

[2]

(b) Write down the number of degrees of freedom for this table.

[1]

(c) Calculate the χ^2 test statistic for this data.

[2]

(d) What conclusion can be drawn from this test? Give a reason for your answer.

[2]

b) Degrees of freedom formula

$df = (c-1)(r-1)$ (not in formula booklet)

where c is the no. of columns and r is the no. of rows.

$c = 4 \quad r = 2$

$df = (4-1)(2-1)$

$df = 3$

At a school in Copenhagen, it is believed that favourite music genre is related to gender.
400 students were asked to indicate their favourite music genre from a selection and the results are indicated in the table below.

	Pop	Rock	Classical	Rap
Female	58	63	17	44
Male	23	96	12	87

It is decided to test this hypothesis by using a χ^2 test at the 5% significance level.
The critical value is 7.815.

(a) State the null and alternative hypotheses for this test.

[2]

(b) Write down the number of degrees of freedom for this table.

[1]

(c) Calculate the χ^2 test statistic for this data.

[2]

(d) What conclusion can be drawn from this test? Give a reason for your answer.

[2]

c) Create a 2x4 matrix on your GDC and perform a χ^2 2-way test.

$\begin{pmatrix} 58 & 63 & 17 & 44 \\ 23 & 96 & 12 & 87 \end{pmatrix}$

$\chi^2 = 33.9844$

$\chi^2 = 34.0$ (3sf)

d) $\chi^2 = 34.0$ $\quad \chi^2_{cv} = 7.815$

$\chi^2 > \chi^2_{cv}$

Reject H_0, there is sufficient evidence to suggest that favourite music genre is dependent on gender.

At a school in Copenhagen, it is believed that favourite music genre is related to gender.
400 students were asked to indicate their favourite music genre from a selection and the results are indicated in the table below.

	Pop	Rock	Classical	Rap
Female	58	63	17	44
Male	23	96	12	87

It is decided to test this hypothesis by using a χ^2 test at the 5% significance level.
The critical value is 7.815.

(a) State the null and alternative hypotheses for this test.

[2]

(b) Write down the number of degrees of freedom for this table.

[1]

(c) Calculate the χ^2 test statistic for this data.

$\chi^2 = 34.0$ (3sf)

[2]

(d) What conclusion can be drawn from this test? Give a reason for your answer.

[2]

Question 2

An environmental organisation is trying to establish if altitude affects the growth of pine needles. A number of needles have been taken from trees at both high and low altitudes and their lengths, in inches, recorded. The results are shown in the table below.

Low altitude	6.1	8.2	7.7	8.0	11.9	6.9	7.5	7.1	8.1
High altitude	7.4	7.9	8.3	6.6	9.5	7.9	8.2	8.1	8.5

Perform a *t*-test to compare the mean lengths of the pine needles.

(a) Write down the null and alternative hypotheses.

[2]

(b) State whether this is a one-tailed test or a two-tailed test.

[1]

(c) Perform a *t*-test at the 10% significance level. Write down the *p*-value.

[2]

(d) Write down the conclusion of the test. Give a reason for your answer.

[2]

a) $H_0: \mu_1 - \mu_2 = 0$

The mean length of the pine needles at low altitude, μ_1, is the same as the mean of the pine needles at high altitude, μ_2.

$H_1: \mu_1 - \mu_2 \neq 0$

The mean length of the pine needles at low altitude, μ_1, is not the same as the mean of the pine needles at high altitude, μ_2.

An environmental organisation is trying to establish if altitude affects the growth of pine needles. A number of needles have been taken from trees at both high and low altitudes and their lengths, in inches, recorded. The results are shown in the table below.

Low altitude	6.1	8.2	7.7	8.0	11.9	6.9	7.5	7.1	8.1
High altitude	7.4	7.9	8.3	6.6	9.5	7.9	8.2	8.1	8.5

Perform a *t*-test to compare the mean lengths of the pine needles.

(a) Write down the null and alternative hypotheses.

$H_0: \mu_1 - \mu_2 = 0 \quad H_1: \mu_1 - \mu_2 \neq 0$

[2]

(b) State whether this is a one-tailed test or a two-tailed test.

[1]

(c) Perform a *t*-test at the 10% significance level. Write down the *p*-valu

[2]

(d) Write down the conclusion of the test. Give a reason for your answe

[2]

b) One-tailed test when

$H_0: \mu_1 - \mu_2 = 0$

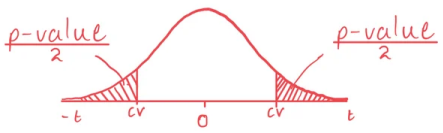
$H_1: \mu_1 - \mu_2 > 0$ or $\mu_1 - \mu_2 < 0$

Two-tailed test when

$H_0: \mu_1 - \mu_2 = 0$

$H_1: \mu_1 - \mu_2 \neq 0$

$H_0: \mu_1 - \mu_2 = 0 \quad H_1: \mu_1 - \mu_2 \neq 0$



∴ This is a two-tailed test.

An environmental organisation is trying to establish if altitude affects the growth of pine needles. A number of needles have been taken from trees at both high and low altitudes and their lengths, in inches, recorded. The results are shown in the table below.

Low altitude	6.1	8.2	7.7	8.0	11.9	6.9	7.5	7.1	8.1
High altitude	7.4	7.9	8.3	6.6	9.5	7.9	8.2	8.1	8.5

Perform a *t*-test to compare the mean lengths of the pine needles.

(a) Write down the null and alternative hypotheses.

$H_0: \mu_1 - \mu_2 = 0 \quad H_1: \mu_1 - \mu_2 \neq 0$

[2]

(b) State whether this is a one-tailed test or a two-tailed test.

∴ This is a two-tailed test.

[1]

(c) Perform a *t*-test at the 10% significance level. Write down the *p*-value.

[2]

(d) Write down the conclusion of the test. Give a reason for your answer.

[2]

c) Input the data into your GDC and perform a 2-Sample t-test.

List 1: Low altitude

List 2: High altitude

p-value = 0.8711...

p-value = 0.871 (3sf)

d) $p = 0.871 \quad SL = 0.1$

$p > SL$

Do not reject H_0, there is insufficient evidence to suggest that the means are different.

An environmental organisation is trying to establish if altitude affects the growth of pine needles. A number of needles have been taken from trees at both high and low altitudes and their lengths, in inches, recorded. The results are shown in the table below.

Low altitude	6.1	8.2	7.7	8.0	11.9	6.9	7.5	7.1	8.1
High altitude	7.4	7.9	8.3	6.6	9.5	7.9	8.2	8.1	8.5

Perform a *t*-test to compare the mean lengths of the pine needles.

(a) Write down the null and alternative hypotheses.

[2]

(b) State whether this is a one-tailed test or a two-tailed test.

[1]

(c) Perform a *t*-test at the 10% significance level. Write down the *p*-value.

[2]

(d) Write down the conclusion of the test. Give a reason for your answer.

[2]

Question 3

A carpet salesman in interested how his sales are distributed and records his sales results over a period of six months. The data is shown in the table.

Month	January	February	March	April	May	June
Number of sales	16	12	14	20	15	19

A chi-squared goodness of fit test is to be performed on the data at the 5% significance level to find out whether the data fits a uniform distribution.

(a) Find an estimate of how many carpets the salesman expects to sell each month.

[1]

(b) Write down the null and alternative hypotheses.

[2]

(c) Write down the number of degrees of freedom for this test.

[1]

(d) Calculate the *p*-value.

[2]

(e) State the conclusion of the test. Give a reason for your answer.

[2]

a) Expected value formula

$E(X) = \frac{\sum_{i=1}^{k} f_i x_i}{n}$ and $n = \sum_{i=1}^{k} f_i$ (in formula booklet)

$E(X) = \frac{16 + 12 + 14 + 20 + 15 + 19}{6}$

$E(X) = 16$ carpets per month

A carpet salesman in interested how his sales are distributed and records his sales results over a period of six months. The data is shown in the table.

Month	January	February	March	April	May	June
Number of sales	16	12	14	20	15	19

A chi-squared goodness of fit test is to be performed on the data at the 5% significance level to find out whether the data fits a uniform distribution.

(a) Find an estimate of how many carpets the salesman expects to sell each month.

[1]

(b) Write down the null and alternative hypotheses.

[2]

(c) Write down the number of degrees of freedom for this test.

[1]

(d) Calculate the *p*-value.

[2]

(e) State the conclusion of the test. Give a reason for your answer.

[2]

b) H_0: The data has a uniform distribution.

H_1: The data does not have a uniform distribution.

A carpet salesman in interested how his sales are distributed and records his sales results over a period of six months. The data is shown in the table.

Month	January	February	March	April	May	June
Number of sales	16	12	14	20	15	19

A chi-squared goodness of fit test is to be performed on the data at the 5% significance level to find out whether the data fits a uniform distribution.

(a) Find an estimate of how many carpets the salesman expects to sell each month.

[1]

(b) Write down the null and alternative hypotheses.

[2]

(c) Write down the number of degrees of freedom for this test.

[1]

(d) Calculate the p-value.

[2]

(e) State the conclusion of the test. Give a reason for your answer.

[2]

c) Degrees of freedom formula

$df = n - 1$ (not in formula booklet)

$df = 6 - 1$

$df = 5$

A carpet salesman in interested how his sales are distributed and records his sales results over a period of six months. The data is shown in the table.

Month	January	February	March	April	May	June
Number of sales	16	12	14	20	15	19

A chi-squared goodness of fit test is to be performed on the data at the 5% significance level to find out whether the data fits a uniform distribution.

(a) Find an estimate of how many carpets the salesman expects to sell each month.

$\bar{x} = 16$ carpets per month

[1]

(b) Write down the null and alternative hypotheses.

[2]

(c) Write down the number of degrees of freedom for this test.

$df = 5$

[1]

(d) Calculate the p-value.

[2]

(e) State the conclusion of the test. Give a reason for your answer.

[2]

d) Input the data into your GDC and perform a χ^2 GOF test.

List 1: Observed list

16, 12, 14, 20, 15, 19

List 2: Expected list

16, 16, 16, 16, 16, 16

$df = 5$

p-value = 0.7192...

p-value = 0.719 (3sf)

A carpet salesman in interested how his sales are distributed and records his sales results over a period of six months. The data is shown in the table.

Month	January	February	March	April	May	June
Number of sales	16	12	14	20	15	19

A chi-squared goodness of fit test is to be performed on the data at the 5% significance level to find out whether the data fits a uniform distribution.

(a) Find an estimate of how many carpets the salesman expects to sell each month.

[1]

(b) Write down the null and alternative hypotheses.

[2]

(c) Write down the number of degrees of freedom for this test.

[1]

(d) Calculate the p-value.

p-value = 0.719 (3sf)

[2]

(e) State the conclusion of the test. Give a reason for your answer.

[2]

e) $p = 0.719$ $SL = 0.05$

$p > SL$

Do not reject H_0, there is insufficient evidence to suggest that the number of carpets sold each month is not uniformly distributed.

Question 4

A supermarket is interested in how the applications for its loyalty scheme are distributed throughout the working week. It is expected that the distribution of the data will be uniform. Over the course of one week the number of applications has been collected and recorded in the table.

Month	Monday	Tuesday	Wednesday	Thursday	Friday
Number of sales	473	405	512	467	503

A goodness of fit test at the 10% significance level is to be performed.
The critical value is 7.779.

(a) Calculate the expected value of the number of applications on any given workday.

[1]

(b) State the null and alternative hypotheses.

[2]

(c) Write down the

(i) χ^2 statistic

(ii) p-value.

[3]

(d) State whether the data fits a uniform distribution, giving a reason for your answer.

[2]

a) Expected value formula

$E(x) = \frac{\sum_{i=1}^{k} f_i x_i}{n}$ and $n = \sum_{i=1}^{k} f_i$ (in formula booklet)

$E(x) = \frac{473 + 405 + 512 + 467 + 503}{5}$

$E(x) = 472$ applications

A supermarket is interested in how the applications for its loyalty scheme are distributed throughout the working week. It is expected that the distribution of the data will be uniform. Over the course of one week the number of applications has been collected and recorded in the table.

Month	Monday	Tuesday	Wednesday	Thursday	Friday
Number of sales	473	405	512	467	503

A goodness of fit test at the 10% significance level is to be performed.
The critical value is 7.779.

(a) Calculate the expected value of the number of applications on any given workday.

[1]

(b) State the null and alternative hypotheses.

[2]

(c) Write down the

(i) χ^2 statistic

(ii) p-value.

[3]

(d) State whether the data fits a uniform distribution, giving a reason for your answer.

[2]

b) H_0 : The data has a uniform distribution.
H_1 : The data does not have a uniform distribution.

A supermarket is interested in how the applications for its loyalty scheme are distributed throughout the working week. It is expected that the distribution of the data will be uniform. Over the course of one week the number of applications has been collected and recorded in the table.

Month	Monday	Tuesday	Wednesday	Thursday	Friday
Number of sales	473	405	512	467	503

A goodness of fit test at the 10% significance level is to be performed.
The critical value is 7.779.

(a) Calculate the expected value of the number of applications on any given workday.

$E(x) = 472$ applications

[1]

(b) State the null and alternative hypotheses.

[2]

(c) Write down the

(i) χ^2 statistic

(ii) p-value.

[3]

(d) State whether the data fits a uniform distribution, giving a reason for your answer.

[2]

c) Input the data into your GDC and perform a χ^2 GOF test.

List 1: Observed list

473, 405, 512, 467, 503

List 2: Expected list

472, 472, 472, 472, 472,

$df = 5 - 1 = 4$

p-value = 0.004718...

$\chi^2 = 14.9915...$

i) $\chi^2 = 15.0$ (3sf)

ii) p-value = 0.00472 (3sf)

A supermarket is interested in how the applications for its loyalty scheme are distributed throughout the working week. It is expected that the distribution of the data will be uniform. Over the course of one week the number of applications has been collected and recorded in the table.

Month	Monday	Tuesday	Wednesday	Thursday	Friday
Number of sales	473	405	512	467	503

A goodness of fit test at the 10% significance level is to be performed.
The critical value is 7.779.

(a) Calculate the expected value of the number of applications on any given workday.

[1]

(b) State the null and alternative hypotheses.

[2]

(c) Write down the

(i) χ^2 statistic $\chi^2 = 15.0$ (3sf)

(ii) p-value. p-value $= 0.00472$ (3sf)

[3]

(d) State whether the data fits a uniform distribution, giving a reason for your answer.

[2]

d) $p = 0.00472$ $SL = 0.1$

$\chi^2 = 15.0$ $\chi^2_{cv} = 7.779$

$p < SL$ and $\chi^2 > \chi^2_{cv}$

$\therefore$ Reject H_0, the data does not fit a uniform distribution.

Question 5

A chi-squared test is performed to see if there is any dependence between eye colour (blue, green or brown) and hair colour (blond, red, brown or black). The test is completed at a significance level of 10%.

(a) State the null and alternative hypotheses.

[2]

(b) Find the number of degrees of freedom for this test.

[2]

(c) The p-value for this test is 0.0726. State the conclusion that can be drawn and justify your answer.

[2]

a) H_0: Hair colour is independent of eye colour.

H_1: Hair colour is not independent of eye colour.

A chi-squared test is performed to see if there is any dependence between eye colour (blue, green or brown) and hair colour (blond, red, brown or black). The test is completed at a significance level of 10%.

(a) State the null and alternative hypotheses.

[2]

(b) Find the number of degrees of freedom for this test.

[2]

(c) The *p*-value for this test is 0.0726. State the conclusion that can be drawn and justify your answer.

[2]

b) Degrees of freedom formula

$df = (c-1)(r-1)$ (not in formula booklet)

where c is the no. of columns and r is the no. of rows.

$c = 3 \quad r = 4$

$df = (3-1)(4-1)$

$df = 6$

A chi-squared test is performed to see if there is any dependence between eye colour (blue, green or brown) and hair colour (blond, red, brown or black). The test is completed at a significance level of 10%.

(a) State the null and alternative hypotheses.

[2]

(b) Find the number of degrees of freedom for this test.

[2]

(c) The *p*-value for this test is 0.0726. State the conclusion that can be drawn and justify your answer.

[2]

c) $p = 0.0726 \quad SL = 0.1$

$p < SL$

$\therefore$ Reject H_0, there is sufficient evidence to suggest that hair colour is not independent of eye colour.

Question 6

It is claimed that women from Japan are taller on average than women from India. The heights, in cm, of 11 women from each country have been collated in the table below.

Japan	India
173.0	155.2
158.2	157.8
148.5	156.0
150.6	142.7
168.7	149.6
149.8	150.1
158.8	152.6
155.3	148.2
159.2	151.3
158.9	147.6
166.0	168.0

(a) Write down the type of test that can be used to compare the means of two sets of data.

[2]

(b) State the null and alternative hypotheses.

[2]

(c) Perform the appropriate test at the 5% significance level.

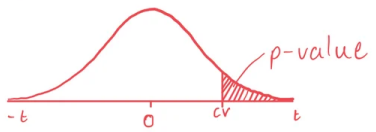
[2]

(d) State whether or not the initial claim is justified. Give a reason for your answer.

[2]

a) "... women from Japan are *taller on average than woman from India."

$\therefore H_0: \mu_1 - \mu_2 = 0 \quad H_1: \mu_1 - \mu_2 > 0$



where μ_1 = average height of Japanese women and μ_2 = average height of Indian women.

∴ This is a one-tailed t-test.

*This would be a two-tailed t-test if the claim was that the average heights of the women were different.

It is claimed that women from Japan are taller on average than women from India. The heights, in cm, of 11 women from each country have been collated in the table below.

Japan	India
173.0	155.2
158.2	157.8
148.5	156.0
150.6	142.7
168.7	149.6
149.8	150.1
158.8	152.6
155.3	148.2
159.2	151.3
158.9	147.6
166.0	168.0

(a) Write down the type of test that can be used to compare the means of two sets of data.

[2]

(b) State the null and alternative hypotheses.

[2]

(c) Perform the appropriate test at the 5% significance level.

[2]

(d) State whether or not the initial claim is justified. Give a reason for your answer.

[2]

b) $H_0: \mu_1 - \mu_2 = 0$

The mean height of women from Japan, μ_1, is the same as the mean height of women from India, μ_2.

$H_1: \mu_1 - \mu_2 > 0$

The mean height of women from Japan, μ_1, greater than the mean height of women from India, μ_2.

It is claimed that women from Japan are taller on average than women from India. The heights, in cm, of 11 women from each country have been collated in the table below.

Japan	India
173.0	155.2
158.2	157.8
148.5	156.0
150.6	142.7
168.7	149.6
149.8	150.1
158.8	152.6
155.3	148.2
159.2	151.3
158.9	147.6
166.0	168.0

(a) Write down the type of test that can be used to compare the means of two sets of data.

[2]

(b) State the null and alternative hypotheses.

$H_0: \mu_1 - \mu_2 = 0 \quad H_1: \mu_1 - \mu_2 > 0$

[2]

(c) Perform the appropriate test at the 5% significance level.

[2]

(d) State whether or not the initial claim is justified. Give a reason for your answer.

[2]

c) Input the data into your GDC and perform a 2-Sample t-test.

List 1: Japan

List 2: India

p-value = 0.03033135...

p-value = 0.0303 (3sf)

d) $p = 0.0303 \quad SL = 0.05$

$p < SL$

∴ Reject H_0, there is sufficient evidence to suggest that the mean height of Japanese women is greater than the mean height of Indian women.

It is claimed that women from Japan are taller on average than women from India. The heights, in cm, of 11 women from each country have been collated in the table below.

Japan	India
173.0	155.2
158.2	157.8
148.5	156.0
150.6	142.7
168.7	149.6
149.8	150.1
158.8	152.6
155.3	148.2
159.2	151.3
158.9	147.6
166.0	168.0

(a) Write down the type of test that can be used to compare the means of two sets of data.

[2]

(b) State the null and alternative hypotheses.

[2]

(c) Perform the appropriate test at the 5% significance level.

p-value = 0.0303 (3sf)

[2]

(d) State whether or not the initial claim is justified. Give a reason for your answer.

[2]

Question 7

The average weight of a newborn baby born at 38 weeks is expected to be less than the average weight of a newborn born at full term (40 weeks). The weights of several babies, in kg, born at 38 weeks and 40 weeks in one hospital are recorded.

38 weeks	3.12	2.87	3.53	3.08	2.86	3.15	3.03	2.99		
40 weeks	3.08	3.59	3.49	3.61	2.99	3.58	3.42	3.55	3.66	3.58

A *t*-test is to be performed at a significance level of 10%.

(a) State the null and alternative hypotheses.

[2]

(b) State whether a one-tailed or a two-tailed test should be used.

[1]

(c) Calculate the *p*-value statistic.

[2]

(d) State whether the initial expectation is confirmed by the test result. Justify your answer.

[2]

a) $H_0: \mu_1 - \mu_2 = 0$

The mean weight of a newborn at 38 weeks, μ_1, is the same as the mean weight of a newborn at 40 weeks, μ_2.

$H_1: \mu_1 - \mu_2 < 0$

The mean weight of a newborn at 38 weeks, μ_1, is less than the mean weight of a newborn at 40 weeks, μ_2.

The average weight of a newborn baby born at 38 weeks is expected to be less than the average weight of a newborn born at full term (40 weeks). The weights of several babies, in kg, born at 38 weeks and 40 weeks in one hospital are recorded.

38 weeks	3.12	2.87	3.53	3.08	2.86	3.15	3.03	2.99		
40 weeks	3.08	3.59	3.49	3.61	2.99	3.58	3.42	3.55	3.66	3.58

A *t*-test is to be performed at a significance level of 10%.

(a) State the null and alternative hypotheses.

[2]

(b) State whether a one-tailed or a two-tailed test should be used.

[1]

(c) Calculate the *p*-value statistic.

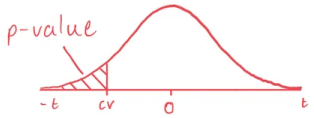
[2]

(d) State whether the initial expectation is confirmed by the test result. Justify your answer.

[2]

b) "... average weight of a newborn baby born at 38 weeks is expected to be less than the average weight of a newborn baby born at full term (40 weeks)"

$\therefore H_0: \mu_1 - \mu_2 = 0 \quad H_1: \mu_1 - \mu_2 < 0$



where μ_1 = average weight of a 38 week newborn and μ_2 = average weight of a 40 week newborn.

$\therefore$ This is a one-tailed t-test.

The average weight of a newborn baby born at 38 weeks is expected to be less than the average weight of a newborn born at full term (40 weeks). The weights of several babies, in kg, born at 38 weeks and 40 weeks in one hospital are recorded.

38 weeks	3.12	2.87	3.53	3.08	2.86	3.15	3.03	2.99		
40 weeks	3.08	3.59	3.49	3.61	2.99	3.58	3.42	3.55	3.66	3.58

A *t*-test is to be performed at a significance level of 10%.

(a) State the null and alternative hypotheses.

[2]

(b) State whether a one-tailed or a two-tailed test should be used.

[1]

(c) Calculate the *p*-value statistic.

[2]

(d) State whether the initial expectation is confirmed by the test result. Justify your answer.

[2]

c) Input the data into your GDC and perform a 2-Sample t-test.

List 1: 38 weeks

List 2: 40 weeks

p-value = 0.001245...

p-value = 0.00125 (3sf)

The average weight of a newborn baby born at 38 weeks is expected to be less than the average weight of a newborn born at full term (40 weeks). The weights of several babies, in kg, born at 38 weeks and 40 weeks in one hospital are recorded.

38 weeks	3.12	2.87	3.53	3.08	2.86	3.15	3.03	2.99		
40 weeks	3.08	3.59	3.49	3.61	2.99	3.58	3.42	3.55	3.66	3.58

A *t*-test is to be performed at a significance level of 10%.

(a) State the null and alternative hypotheses.

[2]

(b) State whether a one-tailed or a two-tailed test should be used.

[1]

(c) Calculate the *p*-value statistic.

[2]

(d) State whether the initial expectation is confirmed by the test result. Justify your answer.

[2]

d) $p = 0.00125$ $SL = 0.1$

$p < SL$

∴ Reject H_0, there is sufficient evidence to suggest that the mean weight of a newborn baby born at 38 weeks is less than the mean weight of a newborn baby born at 40 weeks.

Question 8

A survey was conducted to establish whether particular colours were favoured by people in different age groups. A group of 280 children, teenagers and adults were asked to pick their favourite colour from a choice of red, yellow, blue, green and pink. The observed values are recorded in the table below.

	Red	Yellow	Blue	Green	Pink	Total
Children	20	11	18	8	15	72
Teenagers	22	14	23	20	6	85
Adults	26	21	30	26	20	123
Total	68	46	71	54	41	280

A chi-squared test is to be performed at the 5% significance level. The critical value for the test is 15.507.

(a) Complete the contingency table for the expected values below.

	Red	Yellow	Blue	Green	Pink	Total
Children	17.5	11.8	18.3	13.9	10.5	72
Teenagers	20.6	14.0	21.6	16.4	12.4	85
Adults	29.9	20.2	31.2	23.7	18.0	123
Total	68	46	71	54	41	280

[3]

(b) State the null and alternative hypotheses.

[2]

(c) Write down the number of degrees of freedom.

[1]

(d) Write down the chi-squared test statistic for this data.

[2]

(e) Comment on your results within the context of the question.

[2]

a) Expected values for contingency tables.

$E(x) = \frac{\text{row total}}{\text{total}} \times \frac{\text{column total}}{\text{total}} \times \text{total}$

For example:

Expected value of children who prefer red.

$E(x) = \frac{72}{280} \times \frac{68}{280} \times 280$

$E(x) = 17.5$

A survey was conducted to establish whether particular colours were favoured by people in different age groups. A group of 280 children, teenagers and adults were asked to pick their favourite colour from a choice of red, yellow, blue, green and pink. The observed values are recorded in the table below.

	Red	Yellow	Blue	Green	Pink	Total
Children	20	11	18	8	15	72
Teenagers	22	14	23	20	6	85
Adults	26	21	30	26	20	123
Total	68	46	71	54	41	280

A chi-squared test is to be performed at the 5% significance level. The critical value for the test is 15.507.

(a) Complete the contingency table for the expected values below.

	Red	Yellow	Blue	Green	Pink	Total
Children						
Teenagers						
Adults						
Total						

[3]

(b) State the null and alternative hypotheses.

[2]

(c) Write down the number of degrees of freedom.

[1]

(d) Write down the chi-squared test statistic for this data.

[2]

(e) Comment on your results within the context of the question.

[2]

b)

H_0 : Favourite colour is independent of age.

H_1 : Favourite colour is not independent of age.

A survey was conducted to establish whether particular colours were favoured by people in different age groups. A group of 280 children, teenagers and adults were asked to pick their favourite colour from a choice of red, yellow, blue, green and pink. The observed values are recorded in the table below.

	Red	Yellow	Blue	Green	Pink	Total
Children	20	11	18	8	15	72
Teenagers	22	14	23	20	6	85
Adults	26	21	30	26	20	123
Total	68	46	71	54	41	280

A chi-squared test is to be performed at the 5% significance level. The critical value for the test is 15.507.

(a) Complete the contingency table for the expected values below.

	Red	Yellow	Blue	Green	Pink	Total
Children						
Teenagers						
Adults						
Total						

[3]

(b) State the null and alternative hypotheses.

[2]

(c) Write down the number of degrees of freedom.

[1]

(d) Write down the chi-squared test statistic for this data.

[2]

(e) Comment on your results within the context of the question.

[2]

c) Degrees of freedom formula

$df = (c-1)(r-1)$ (not in formula booklet)

where c is the no. of columns and r is the no. of rows.

$c = 5 \quad r = 3$

$df = (5-1)(3-1)$

$df = 8$

A survey was conducted to establish whether particular colours were favoured by people in different age groups. A group of 280 children, teenagers and adults were asked to pick their favourite colour from a choice of red, yellow, blue, green and pink. The observed values are recorded in the table below.

	Red	Yellow	Blue	Green	Pink	Total
Children	20	11	18	8	15	72
Teenagers	22	14	23	20	6	85
Adults	26	21	30	26	20	123
Total	68	46	71	54	41	280

A chi-squared test is to be performed at the 5% significance level. The critical value for the test is 15.507.

(a) Complete the contingency table for the expected values below.

	Red	Yellow	Blue	Green	Pink	Total
Children						
Teenagers						
Adults						
Total						

[3]

(b) State the null and alternative hypotheses.

[2]

(c) Write down the number of degrees of freedom.

[1]

(d) Write down the chi-squared test statistic for this data.

[2]

(e) Comment on your results within the context of the question.

[2]

d) Create a 3×5 matrix on your GDC and perform a χ^2 2-way test.

$$\begin{pmatrix} 20 & 11 & 18 & 8 & 15 \\ 22 & 14 & 23 & 20 & 6 \\ 26 & 21 & 30 & 26 & 20 \end{pmatrix}$$

$\chi^2 = 10.138$

$\chi^2 = 10.1$ (3sf)

A survey was conducted to establish whether particular colours were favoured by people in different age groups. A group of 280 children, teenagers and adults were asked to pick their favourite colour from a choice of red, yellow, blue, green and pink. The observed values are recorded in the table below.

	Red	Yellow	Blue	Green	Pink	Total
Children	20	11	18	8	15	72
Teenagers	22	14	23	20	6	85
Adults	26	21	30	26	20	123
Total	68	46	71	54	41	280

A chi-squared test is to be performed at the 5% significance level. The critical value for the test is 15.507.

(a) Complete the contingency table for the expected values below.

	Red	Yellow	Blue	Green	Pink	Total
Children						
Teenagers						
Adults						
Total						

[3]

(b) State the null and alternative hypotheses.

[2]

(c) Write down the number of degrees of freedom.

[1]

(d) Write down the chi-squared test statistic for this data.

$\chi^2 = 10.1$ (3sf)

[2]

(e) Comment on your results within the context of the question.

[2]

e) $\chi^2 = 10.138$ $\chi^2_{cv} = 15.507$

$\chi^2 < \chi^2_{cv}$

∴ Do not reject H_0, there is insufficient evidence to suggest that the choice of favourite colour is dependent of age.

Question 9

The heights of giraffes are normally distributed with a mean of 3.8 m and a standard deviation of 0.7 m. As part of a conservation project in Kenya, the heights of 350 giraffes are measured and the results of the survey are seen in the table below. A chi-squared test at a significance level of 10% is to be performed to determine if the surveyed giraffes fit the normal distribution stated.

Height (cm)	Frequency
$h < 3$	50
$3 \leq h < 4$	160
$4 \leq h < 5$	119
$5 \leq h < 6$	21

(a) Complete the following table of expected heights.

Height (cm)	Probability	Expected frequency
$h < 3$	0.126549	44.3
$3 \leq h < 4$	0.485902	170.1
$4 \leq h < 5$	0.34431	120.5
$5 \leq h < 6$	0.042401	14.8

[3]

(b) Write down the null and alternative hypotheses.

[2]

(c) State the number of degrees of freedom.

[1]

(d) Calculate the *p*-value.

[2]

(e) State whether the results of the chi-squared test support the null hypothesis. Justify your answer.

[2]

a) Calculate the probabilities using the normal CDF function on your GDC.

Calculate the expected frequencies by multiplying the probabilities by 350.

The heights of giraffes are normally distributed with a mean of 3.8 m and a standard deviation of 0.7 m. As part of a conservation project in Kenya, the heights of 350 giraffes are measured and the results of the survey are seen in the table below. A chi-squared test at a significance level of 10% is to be performed to determine if the surveyed giraffes fit the normal distribution stated.

Height (cm)	Frequency
$h < 3$	50
$3 \leq h < 4$	160
$4 \leq h < 5$	119
$5 \leq h < 6$	21

(a) Complete the following table of expected heights.

Height (cm)	Probability	Expected frequency
$h < 3$		
$3 \leq h < 4$		
$4 \leq h < 5$		
$5 \leq h < 6$		

[3]

(b) Write down the null and alternative hypotheses.

[2]

(c) State the number of degrees of freedom.

[1]

(d) Calculate the *p*-value.

[2]

(e) State whether the results of the chi-squared test support the null hypothesis. Justify your answer.

[2]

b) H_0: The heights of the giraffes fit a normal distribution.

H_1: The heights of the giraffes do not fit a normal distribution.

The heights of giraffes are normally distributed with a mean of 3.8 m and a standard deviation of 0.7 m. As part of a conservation project in Kenya, the heights of 350 giraffes are measured and the results of the survey are seen in the table below. A chi-squared test at a significance level of 10% is to be performed to determine if the surveyed giraffes fit the normal distribution stated.

4

Height (cm)	Frequency
$h < 3$	50
$3 \leq h < 4$	160
$4 \leq h < 5$	119
$5 \leq h < 6$	21

(a) Complete the following table of expected heights.

Height (cm)	Probability	Expected frequency
$h < 3$		
$3 \leq h < 4$		
$4 \leq h < 5$		
$5 \leq h < 6$		

[3]

(b) Write down the null and alternative hypotheses.

[2]

(c) State the number of degrees of freedom.

[1]

(d) Calculate the *p*-value.

[2]

(e) State whether the results of the chi-squared test support the null hypothesis. Justify your answer.

[2]

c) Degrees of freedom formula (not in formula booklet)

$df = n - 1$

$df = 4 - 1$

$df = 3$

The heights of giraffes are normally distributed with a mean of 3.8 m and a standard deviation of 0.7 m. As part of a conservation project in Kenya, the heights of 350 giraffes are measured and the results of the survey are seen in the table below. A chi-squared test at a significance level of 10% is to be performed to determine if the surveyed giraffes fit the normal distribution stated.

Height (cm)	Frequency
$h < 3$	50
$3 \leq h < 4$	160
$4 \leq h < 5$	119
$5 \leq h < 6$	21

(a) Complete the following table of expected heights.

Height (cm)	Probability	Expected frequency
$h < 3$	0.126549	44.3
$3 \leq h < 4$	0.485902	170.1
$4 \leq h < 5$	0.34431	120.5
$5 \leq h < 6$	0.042401	14.8

[3]

(b) Write down the null and alternative hypotheses.

[2]

(c) State the number of degrees of freedom.

df = 3

[1]

(d) Calculate the p-value.

[2]

(e) State whether the results of the chi-squared test support the null hypothesis. Justify your answer.

[2]

d) Input the data into your GDC and perform a X^2 GOF test.

List 1: Observed list

50, 160, 119, 21

List 2: Expected list

44.3, 170.1, 120.5, 14.8

df = 3

p-value = 0.26701...

p-value = 0.267 (3sf)

The heights of giraffes are normally distributed with a mean of 3.8 m and a standard deviation of 0.7 m. As part of a conservation project in Kenya, the heights of 350 giraffes are measured and the results of the survey are seen in the table below. A chi-squared test at a significance level of 10% is to be performed to determine if the surveyed giraffes fit the normal distribution stated.

Height (cm)	Frequency
$h < 3$	50
$3 \leq h < 4$	160
$4 \leq h < 5$	119
$5 \leq h < 6$	21

(a) Complete the following table of expected heights.

Height (cm)	Probability	Expected frequency
$h < 3$		
$3 \leq h < 4$		
$4 \leq h < 5$		
$5 \leq h < 6$		

[3]

(b) Write down the null and alternative hypotheses.

[2]

(c) State the number of degrees of freedom.

[1]

(d) Calculate the p-value.

p-value = 0.267 (3sf)

[2]

(e) State whether the results of the chi-squared test support the null hypothesis. Justify your answer.

[2]

e) p = 0.267 SL = 0.1

p > SL

∴ Do not reject Ho, there is not sufficient evidence to suggest that the heights of these giraffes does not fit a normal distribution.

Question 10

Juan undertakes an investigation into reaction times for his Mathematics IA. As part of the investigation he wants to determine if reaction speed is dependent on gender. He randomly selects 60 people in his year group and records their gender and the average time they take to complete a reaction test.

Juan decides to run a χ^2 test on the data. The results are displayed in the table below.

	Quick $300 \leq R < 450$	Moderate $450 \leq R < 600$	Slow $600 \leq R < 750$	Very Slow $750 \leq R < 900$	Total
Female	6	11	13	9	39
Male	2	6	7	6	21
Total	8	17	20	15	60

(a) State the null and alternative hypotheses.

[2]

(b) Calculate the expected value for the number of male students that have a "Quick" reaction time.

[2]

(c) Explain why the result of performing a χ^2 test on the data, organised as it currently is, would not produce a valid result.

[1]

(d) Comment on what action Juan could take to reorganise the data in order to improve the validity of the result of the χ^2 test.

[1]

a) H_0: Reaction time is independent of gender.

H_1: Reaction time is dependent on gender.

Juan undertakes an investigation into reaction times for his Mathematics IA. As part of the investigation he wants to determine if reaction speed is dependent on gender. He randomly selects 60 people in his year group and records their gender and the average time they take to complete a reaction test.

Juan decides to run a χ^2 test on the data. The results are displayed in the table below.

	Quick $300 \leq R < 450$	Moderate $450 \leq R < 600$	Slow $600 \leq R < 750$	Very Slow $750 \leq R < 900$	Total
Female	6	11	13	9	39
Male	2	6	7	6	21
Total	8	17	20	15	60

(a) State the null and alternative hypotheses.

[2]

(b) Calculate the expected value for the number of male students that have a "Quick" reaction time.

[2]

(c) Explain why the result of performing a χ^2 test on the data, organised as it currently is, would not produce a valid result.

[1]

(d) Comment on what action Juan could take to reorganise the data in order to improve the validity of the result of the χ^2 test.

[1]

b) $n(\text{Quick}) = 8$, $n(\text{Male}) = 21$

$\frac{8}{60} \times \frac{21}{60} \times 60 = 2.8$

2.8

Juan undertakes an investigation into reaction times for his Mathematics IA. As part of the investigation he wants to determine if reaction speed is dependent on gender. He randomly selects 60 people in his year group and records their gender and the average time they take to complete a reaction test.

Juan decides to run a χ^2 test on the data. The results are displayed in the table below.

	Quick $300 \leq R < 450$	Moderate $450 \leq R < 600$	Slow $600 \leq R < 750$	Very Slow $750 \leq R < 900$	Total
Female	6	11	13	9	39
Male	2	6	7	6	21
Total	8	17	20	15	60

(a) State the null and alternative hypotheses.

[2]

(b) Calculate the expected value for the number of male students that have a "Quick" reaction time.

[2]

(c) Explain why the result of performing a χ^2 test on the data, organised as it currently is, would not produce a valid result.

[1]

(d) Comment on what action Juan could take to reorganise the data in order to improve the validity of the result of the χ^2 test.

[1]

c) Expected frequency values must be ≥5 to produce a valid result from a chi-squared test.

Juan undertakes an investigation into reaction times for his Mathematics IA. As part of the investigation he wants to determine if reaction speed is dependent on gender. He randomly selects 60 people in his year group and records their gender and the average time they take to complete a reaction test.

Juan decides to run a χ^2 test on the data. The results are displayed in the table below.

	Quick $300 \leq R < 450$	Moderate $450 \leq R < 600$	Slow $600 \leq R < 750$	Very Slow $750 \leq R < 900$	Total
Female	6	11	13	9	39
Male	2	6	7	6	21
Total	8	17	20	15	60

(a) State the null and alternative hypotheses.

[2]

(b) Calculate the expected value for the number of male students that have a "Quick" reaction time.

[2]

(c) Explain why the result of performing a χ^2 test on the data, organised as it currently is, would not produce a valid result.

[1]

(d) Comment on what action Juan could take to reorganise the data in order to improve the validity of the result of the χ^2 test.

[1]

d) The grouping of numerical data could be reduced to 3 headings: Quick, moderate and slow.

Question 11

An international school of 293 students in Ho Chi Minh City holds a mock election parallel to the UK general election. The age of each voter is recorded on their ballot paper along with their vote for either the Conservative, Labour or Liberal party. The results are listed in the table below.

	Conservative	Labour	Liberal	Total
$a \leq 12$	5	8	9	22
$12 < a \leq 14$	13	33	45	91
$14 < a \leq 16$	17	21	71	109
$16 < a \leq 18$	11	31	30	72
Total	46	93	155	294

A chi-squared test at the 10% significance level is to be performed on the data.

(a) State the null and alternative hypotheses.

[2]

a) H_0: The age of a voter is independent of their vote.

H_1: The age of a voter is not independent of their vote.

The table below shows the expected values for the data.

	Conservative	Labour	Liberal	Total
$a \leq 12$	3.4	7.0	11.6	22
$12 < a \leq 14$	14.2	28.8	48.0	91
$14 < a \leq 16$	17.1	34.5	57.5	109
$16 < a \leq 18$	11.3	22.8	38.0	72
Total	46	93	155	294

(b) (i) Explain why the table should be re-drawn before conducting the test.

(ii) Re-draw the table of observed values by combining the results for $a \leq 12$ with those for $12 < a \leq 14$.

[3]

(c) Calculate the p-value using the re-drawn table from part (b) and state the conclusion for the test, giving a reason for your answer.

[4]

An international school of 293 students in Ho Chi Minh City holds a mock election parallel to the UK general election. The age of each voter is recorded on their ballot paper along with their vote for either the Conservative, Labour or Liberal party. The results are listed in the table below.

	Conservative	Labour	Liberal	Total
$a \leq 12$	5	8	9	22
$12 < a \leq 14$	13	33	45	91
$14 < a \leq 16$	17	21	71	109
$16 < a \leq 18$	11	31	30	72
Total	46	93	155	294

A chi-squared test at the 10% significance level is to be performed on the data.

(a) State the null and alternative hypotheses.

[2]

The table below shows the expected values for the data.

	Conservative	Labour	Liberal	Total
$a \leq 12$	3.4	7.0	11.6	22
$12 < a \leq 14$	14.2	28.8	48.0	91
$14 < a \leq 16$	17.1	34.5	57.5	109
$16 < a \leq 18$	11.3	22.8	38.0	72
Total	46	93	155	294

(b) (i) Explain why the table should be re-drawn before conducting the test.

(ii) Re-draw the table of observed values by combining the results for $a \leq 12$ with those for $12 < a \leq 14$.

[3]

(c) Calculate the p-value using the re-drawn table from part (b) and state the conclusion for the test, giving a reason for your answer.

[4]

b) i) All expected values should be equal to or more than 5.

This a condition for a χ^2 GOF test.

ii)

	Conservative	Labour	Liberal	Total
$a \leq 14$	18	41	54	113
$14 < a \leq 16$	17	21	71	109
$16 < a \leq 18$	11	31	30	72
Total	46	93	155	294

An international school of 293 students in Ho Chi Minh City holds a mock election parallel to the UK general election. The age of each voter is recorded on their ballot paper along with their vote for either the Conservative, Labour or Liberal party. The results are listed in the table below.

	Conservative	Labour	Liberal	Total
$a \leq 12$	5	8	9	22
$12 < a \leq 14$	13	33	45	91
$14 < a \leq 16$	17	21	71	109
$16 < a \leq 18$	11	31	30	72
Total	46	93	155	294

A chi-squared test at the 10% significance level is to be performed on the data.

(a) State the null and alternative hypotheses.

[2]

The table below shows the expected values for the data.

	Conservative	Labour	Liberal	Total
$a \leq 12$	3.4	7.0	11.6	22
$12 < a \leq 14$	14.2	28.8	48.0	91
$14 < a \leq 16$	17.1	34.5	57.5	109
$16 < a \leq 18$	11.3	22.8	38.0	72
Total	46	93	155	294

(b) (i) Explain why the table should be re-drawn before conducting the test.

(ii) Re-draw the table of observed values by combining the results for $a \leq 12$ with those for $12 < a \leq 14$.

[3]

(c) Calculate the p-value using the re-drawn table from part (b) and state the conclusion for the test, giving a reason for your answer.

[4]

c) Create 3 x 3 matrix on your GDC.

$\begin{pmatrix} 18 & 41 & 54 \\ 17 & 21 & 71 \\ 11 & 31 & 30 \end{pmatrix}$ → values taken from part (b) (ii).

Perform a χ^2 2-way test

p-value = 0.006115

p-value = 0.00612 < SL = 0.1

Reject H_0, there is sufficient evidence to suggest that a voter's choice of political party is dependent on their age.