1. A random sample of 100 people were asked if their finances were worse, the same or better than this time last year. The sample was split according to their annual income and the results are shown in the table below.

Finances / Annual income	Worse	Same	Better
Under £15 000	14	11	9
£15 000 and above	17	20	29

Test, at the 5% level of significance, whether or not the relative state of their finances is independent of their income range. State your hypotheses and show your working clearly.

(Total 10 marks)

2. A research worker studying colour preference and the age of a random sample of 50 children obtained the results shown below.

Age in years	Red	Blue	Totals
4	12	6	18
8	10	7	17
12	6	9	15
Totals	28	22	50

Using a 5% significance level, carry out a test to decide whether or not there is an association between age and colour preference. State your hypotheses clearly.

(Total 11 marks)

3. People over the age of 65 are offered an annual flu injection.
A health official took a random sample from a list of patients who were over 65. She recorded their gender and whether or not the offer of an annual flu injection was accepted or rejected. The results are summarised below.

Gender	Accepted	Rejected
Male	170	110
Female	280	140

Using a 5% significance level, test whether or not there is an association between gender and acceptance or rejection of an annual flu injection. State your hypotheses clearly.

(Total 9 marks)

4. A researcher carried out a survey of three treatments for a fruit tree disease. The contingency table below shows the results of a survey of a random sample of 60 diseased trees.

	No action	Remove diseased branches	Spray with chemicals
Tree died within 1 year	10	5	6
Tree survived for 1–4 years	5	9	7
Tree survived beyond 4 years	5	6	7

Test, at the 5% level of significance, whether or not there is any association between the treatment of the trees and their survival. State your hypotheses and conclusion clearly.

(Total 11 marks)

5. A random sample of 500 adults completed a questionnaire on how often they took part in some form of exercise. They gave a response of 'never', 'sometimes' or 'regularly'. Of those asked, 52% were females of whom 10% never exercised and 35% exercised regularly. Of the males, 12.5% never exercised and 55% sometimes exercised.

Test, at the 5% level of significance, whether or not there is any association between gender and the amount of exercise. State your hypotheses clearly.

(Total 12 marks)

6. A new drug to treat the common cold was used with a randomly selected group of 100 volunteers. Each was given the drug and their health was monitored to see if they caught a cold. A randomly selected control group of 100 volunteers was treated with a dummy pill. The results are shown in the table below.

	Cold	No cold
Drug	34	66
Dummy pill	45	55

Using a 5% significance level, test whether or not the chance of catching a cold is affected by taking the new drug. State your hypotheses clearly.

(Total 11 marks)

1.

Finances / Income	Worse	Same	Better	
Under £15 000	10.54	10.54	12.92	34
£15 000 and above	20.46	20.46	25.08	66
	31	31	38	100

M1

A1

H_0 : State of finances and income are independent (not associated)

H_1 : State of finances and income are not independent (associated) B1

O_i	E_i	$\frac{(O_i - E_i)^2}{E_i}$	$\frac{O_i^2}{E_i}$
14	10.54	1.1358….	18.59..
11	10.54	0.0200….	11.48..
9	12.92	1.1893…	6.269..
17	20.46	0.5851…	14.12..
20	20.46	0.0103…	19.55..
29	25.08	0.6126…	33.53..

M1

A1

$\sum \frac{(O_i - E_i)^2}{E_i} = 3.553\ldots$ or $\sum \frac{O_i^2}{E_i} - 100 = 103.553\ldots - 100 = 3.553\ldots$

(awrt **3.55)** A1

$\nu = (3-1)(2-1) = 2$ B1

cv is 5.991 B1

3.553 < 5.991 so insufficient evidence to reject H_0 or not significant M1

There is no evidence of association between state of finances and income. A1

Note

1st M1 for some use of $\frac{\text{Row Total} \times \text{Col.Total}}{\text{Grand Total}}$. May be implied by correct E_i

1st A1 for all expected frequencies correct

B1 for both hypotheses. Must mention "state" or "finances" and "income" at least once Use of "relationship" or "correlation" or "connection" is B0

2nd M1 for at least two correct terms (as in 3rd or 4th column) or correct expressions with their E_i

2nd A1 for all correct terms. May be implied by a correct answer. (2 dp or better-allow eg 1.13…)

3rd M1 for a correct statement linking their test statistic and their cv. Must be χ^2 not normal.

4th A1 for a correct comment in context – must mention "state" or "finances" and "income" condone "relationship" or "connection" here but **not** "correlation". No follow through. e.g. "There is no evidence of a relationship between finances and income"

[10]

2. H_0 : No association between age and colour (Independent) B1

H_1 : Association between age and colour (Not independent) B1

O	E	$\frac{(O-E)^2}{E}$
12	10.08	0.3657…
6	7.92	0.4654…
10	9.52	0.0242…
7	7.48	0.0308…
6	8.4	0.6857…
9	6.6	0.8727…

at least one $\frac{R_T \times C_T}{G_T}$, *B11* M1 A1

$\frac{(O-E)^2}{E}$ M1 A1

3 s.f. or better

$\sum \frac{(O-E)^2}{E} = 2.4446...$ $\sum$, awrt 2.44 M1 A1

$\nu = (3-1)(2-1) = 2, \chi^2_2 = 5.991$ B1 B1ft

Insufficient evidence to reject H_0

No association between age and colour A1ft

[11]

3. H_0: No association between gender and acceptance B1
H_1 : gender and acceptance are associated

	Accept	Not accept	Total
Males	170 (180)	110 (100)	280
Females	280 (270)	140 (150)	420
Totals	450	250	700

Expected Values M1 A1

O	***E***	$\frac{(O-E)^2}{E}$
170	180	0.5556
110	100	1.0000
280	270	0.3704
140	150	0.6667

$\sum \frac{(O-E)^2}{E}$ = 2.59 (Yates' 2.34) (Condone use of Yates') M1 A1

ν = 1; (5%) = 3.841 B1; B1

3.841 > 2.59. There is insufficient evidence to reject Ho M1
There is no association between a persons gender and their acceptance A1
(of the offer of a flu jab.) 9

[9]

4.

	No action	Remove diseased branches	Spray with Chemicals	Totals
Tree died within 1 year	10(7)	5(7)	6(7)	21
Survived 1–4 years	5(7)	9(7)	7(7)	21
Survived > 4 years	5(6)	6(6)	7(6)	18
Totals	20	20	20	60

$\frac{RT \times CT}{GT}$ M1

6×7 A1

3×6 A1

H_0: Treatment & survival are independent (not associated) B1 both

H_1: Treatment & survival are not independent (associated)

$\alpha = 0.05$

$\propto = (3-1) \times (3-1) = 4$ B1

CR: $\chi^2 > 9.488$ B1ft

$$\sum \frac{(O-E)^2}{E} = \frac{9}{7} + \frac{4}{7} + \frac{1}{7} + \frac{4}{7} + \frac{4}{7} + 0 + \frac{1}{6} + 0 + \frac{1}{6}$$

Use of $\sum \frac{(O-E)^2}{E}$ M1

Any 2 values A1

$= 3.47619\ldots$ A1

awrt 3.48

Since 3.47619… is NOT in the critical region (ie < 9.488) there is insufficient evidence to reject H_0.

There is no evidence of association between treatment and length of survival.

Comparison M1

Conclusion A1ft

[11]

5.

	Never	Sometimes	Regularly	Totals
Males	30	132	78	240
Females	26	143	91	260
	56	275	169	500

M1 convert % to freq

A1 (26, 91, 30, 132)

A1 (143, 78)

H_0: No association (independent) between gender and exercise B1

H_1: association (not independent) between gender and exercise B1

Expected Values

	Never	Sometimes	Regularly	Totals
Males	26.88	132	81.12	240
Females	29.12	143	87.88	260
	56	275	169	500

M1

A1 at least 3sf

$\alpha = 0.05$ $\underline{\nu = 2}$; CV $\chi^2 > \underline{5.991}$ B1; B1ft

$\sum \frac{(O-E)^2}{E}$ *OR* $\sum \frac{O^2}{E} - N = 0.9271$ M1 A1

answers in range 0.90 – 0.95

Not in critical region – no evidence of association between gender and exercise A1 ft

[12]

6. H_0: Taking drug and catching a cold are independent (not associated) B1
H_1: Taking drug and catching a cold are not independent (associated)(not ditto)

Both B1
All totals B1
$E = \dfrac{RT \times CT}{GT}$ M1 A1 A1

	Cold	**NoCold**	
Drug	**34(39.5)**	**66(60.5)**	**100**
Dummy	**45(39.5)**	**55(60.5)**	**100**
	79	**121**	**200**

O	E	$\dfrac{(O-E)^2}{E}$
34	**39.5**	**0.766**
66	**60.5**	**0.5**
45	**39.5**	**0.765**
55	**60.5**	**0.5**

$\sum \dfrac{(O-E)^2}{E} = 2.53$ (NB with Yates 2.09)
attempt & add, awrt 0.766 & 0.5 twice, awrt 2.53 M1 A1 A1

$v = 1$, $\chi_1^2(5\%) = 3.841 > 2.53$ 1, 3.841 B1, B1

No reason to believe that the chance of catching a cold is affected by taking the new drug A1∫

[11]

1. For most candidates this question was a good source of marks. Hypotheses were usually correctly phrased in terms of "independence" or "association" and the calculations were usually clearly set out although some inappropriate rounding sometimes gave an answer of 3.56. The degrees of freedom and critical value caused few problems and most gave a correct conclusion in context.

2. There were some excellent responses with a large number of correct answers seen. It was unusual not to see hypotheses well stated and the conclusion given correctly in context.

3. This question was answered well and most candidates scored full or almost full marks. The Expected frequencies were almost always correct as was the calculation of the test statistic. Occasionally the hypotheses were the wrong way around and sometimes the conclusion was not given in context.

4. Many candidates produced completely correct solutions to this question. Ill-defined hypotheses, poor arithmetic and not giving the conclusion in context were the common errors.

5. This question was very well done with even very weak candidates managing to gain 11 or 12 marks. Some candidates tried to use percentages rather than frequencies and a few did not seem able to use the χ^2 table correctly. As is often the case hypotheses were sometimes reversed.

6. There was a sizeable group of candidates who answered this question well as they were able to demonstrate a good grasp of the mathematical technique required. The most common errors involved accuracy in the working, careless definitions of the hypotheses and missing context in the conclusion.