Chi-squared tests 6E

1 A 3 × 2 table has 3 rows (h) and 2 columns (k) so:

Degrees of freedom $\nu = (3-1)(2-1) = 2$

Critical value is $\chi^2_2(5\%) = 5.991$

2 H_0: Ownership is not related to the locality.
H_1: Ownership is related to the locality.

Degrees of freedom $\nu = (3-1)(2-1) = 2$

Critical value is $\chi^2_2(5\%) = 5.991$

Test statistic $X^2 = 13.1$, which is greater than 5.991 so reject H_0 at 5% significance level and conclude that there is evidence to suggest that ownership of a television is related to the locality.

3 **a** There are three groups of students and so three rows of data, and three exam classifications so three columns of data. When calculating the expected values, it is not necessary to calculate the last value in each row because it has to equal the row total. Similarly, it is not necessary to calculate the last value in each column because it has to equal the column total.

So degrees of freedom $= (3-1)\times(3-1) = 4$

b H_0: There is no association between the student group and the results.
H_1: There is an association between the student group and the results.

Critical value is $\chi^2_4(5\%) = 9.488$

Test statistic $X^2 = 10.28$, which is greater than 9.488 so reject H_0 at 5% significance level and conclude that there is evidence to suggest an association between the student groups and the exam results they achieve.

4 H_0: There is no association between Mathematics and English results.
H_1: There is an association between Mathematics and English results.

These are the observed frequencies (O_i) with totals for each row and column:

		English grades			
		A	**B**	**C**	**Total**
Maths grades	**A**	17	28	18	63
	B	38	45	16	99
	C	12	12	14	38
	Total	67	85	48	200

Calculate the expected frequencies (E_i) for each cell. For example:

Expected frequency Mathematics A and English A $= \dfrac{63 \times 67}{200} = 21.105$

The expected frequencies (E_i) are:

		English grades		
		A	**B**	**C**
Maths grades	**A**	21.105	26.775	15.120
	B	33.165	42.075	23.760
	C	12.730	16.150	9.120

The test statistic (X^2) calculations are:

O_i	E_i	$\dfrac{(O_i - E_i)^2}{E_i}$
17	21.105	0.7985
28	26.775	0.0560
18	15.120	0.5486
38	33.165	0.7049
45	42.075	0.2033
16	23.760	2.5344
12	12.730	0.0419
12	16.150	1.0664
14	9.120	2.6112

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 8.565$$

The number of degrees of freedom $\nu = (3-1)(3-1) = 4$; from the tables: $\chi_4^2(5\%) = 9.488$

As 8.565 is less than 9.488, there is insufficient evidence to reject H_0 at the 5% level. This suggest that the Mathematics and English results are independent.

5 H_0: There is no association between station and lateness.
H_1: There is an association between station and lateness.

These are the observed frequencies (O_i) with totals for each row and column:

		On time	Late	Total
Station	***A***	26	14	40
	B	30	10	40
	C	44	26	70
	Total	100	50	150

Calculate the expected frequencies (E_i) for each cell. For example:

Expected frequency 'On time' and 'Station A' $= \frac{40 \times 100}{150} = 26.666...$

The expected frequencies (E_i) are:

		On time	Late
Station	***A***	26.666	13.333
	B	26.666	13.333
	C	46.666	23.333

The test statistic (X^2) calculations are:

O_i	E_i	$\frac{(O_i - E_i)^2}{E_i}$
26	26.666	0.0166
14	13.333	0.0334
30	26.666	0.4168
10	13.333	0.8332
44	46.666	0.1523
26	23.333	0.3048

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 1.757$$

The number of degrees of freedom $\nu = (3-1)(2-1) = 2$; from the tables: $\chi_2^2(5\%) = 5.991$

As 1.757 is less than 5.991, there is insufficient evidence to reject H_0 at the 5% level. There is no reason to believe there is an association between station and lateness.

6 H_0: There is no association between gender and grades.
H_1: There is an association between gender and grades.

The number of degrees of freedom $\nu = (5-1)(2-1) = 4$; from the tables: $\chi_4^2(1\%) = 13.277$

As X^2 (= 14.27) is greater than 13.277, reject H_0 at the 1% level. There is evidence to suggest an association between gender and grade.

7 a The contingency table showing observed frequencies is:

	Factory		
	A	***B***	**Total**
OK	47	28	75
Defective	13	12	25
Total	60	40	100

b H_0: There is no association between factory and quality.
H_1: There is an association between factory and quality.

Calculate the expected frequencies (E_i) for each cell. For example:

Expected frequency 'OK' and 'Factory A' $= \frac{75 \times 60}{100} = 45$

The expected frequencies (E_i) are:

	Factory	
	A	***B***
OK	45	30
Defective	15	10

The test statistic (X^2) calculations are:

O_i	E_i	$\frac{(O_i - E_i)^2}{E_i}$
47	45	0.0889
28	30	0.1333
13	15	0.2667
12	10	0.4

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 0.8889$$

The number of degrees of freedom $\nu = (2-1)(2-1) = 1$; from the tables: $\chi_1^2(5\%) = 3.841$

As 0.8889 is less than 3.841, there is insufficient evidence to reject H_0 at the 5% level. There is no reason to believe there is an association between a factory and garment quality.

8 H_0: There is no association between gender and susceptibility to influenza.
H_1: There is an association between gender and susceptibility to influenza.

The contingency table showing observed frequencies is:

	Boys	**Girls**	**Total**
Flu	15	8	23
No flu	7	20	27
Total	22	28	50

Calculate the expected frequencies (E_i) for each cell. For example:

Expected frequency 'Boy' and 'Flu' $= \dfrac{23 \times 22}{50} = 10.12$

The expected frequencies (E_i) are:

	Boys	**Girls**
Flu	10.12	12.88
No flu	11.88	15.12

The test statistic (X^2) calculations are:

O_i	E_i	$\dfrac{(O_i - E_i)^2}{E_i}$
15	10.12	2.343
8	12.88	1.849
7	11.88	2.005
20	15.12	1.575

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 7.781$$

The number of degrees of freedom $\nu = (2-1)(2-1) = 1$; from the tables: $\chi_1^2(5\%) = 3.841$

As 7.781 is greater than 3.841, reject H_0 at the 5% level. There is evidence of an association between the gender and susceptibility to influenza.

9 H_0: There is no association between choice of beach and the gender of the organisms.
H_1: There is an association between choice of beach and the gender of the organisms.

Observed (O_i)		**Beach**			
		A	***B***	***C***	**Total**
Gender	**Male**	46	80	40	166
	Female	54	120	160	334
	Total	100	200	200	500

Calculate the expected value for each cell by multiplying column and row totals and dividing by the grand total, for example:

$$E_{A,M} = \frac{166 \times 100}{500} = 33.2$$

Expected (E_i)		**Beach**		
		A	***B***	***C***
Gender	**Male**	33.2	66.4	66.4
	Female	66.8	133.6	133.6

The test statistic (X^2) calculations are:

O_i	E_i	$\frac{(O_i - E_i)^2}{E_i}$
46	33.2	0.0166
80	66.4	0.0334
40	66.4	0.4168
54	66.8	0.8332
120	133.6	0.1523
160	133.6	0.3048

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 27.271$$

The number of degrees of freedom $\nu = (3-1)(2-1) = 2$; from the tables: $\chi_2^2(5\%) = 5.991$

As 27.271 is greater than 5.991, reject H_0 at the 5% level. There is evidence of an association between the gender of an organism and the beach on which it is found.

10 H_0: There is no association between age and number of credit cards.
H_1: There is an association between age and number of credit cards.

These are the observed frequencies (O_i) with totals for each row and column:

		Number of cards		
		$\leqslant 3$	> 3	Total
Age	$\leqslant 30$	74	20	94
	> 30	50	35	85
	Total	124	55	179

Calculate the expected frequencies (E_i) for each cell. For example:

Expected frequency '$\leqslant 30$' and '$\leqslant 3$' $= \dfrac{94 \times 124}{179} = 65.117$

The expected frequencies (E_i) are:

		Number of cards	
		$\leqslant 3$	> 3
Age	$\leqslant 30$	65.117	28.883
	> 30	58.883	26.117

The test statistic (X^2) calculations are:

O_i	E_i	$\dfrac{(O_i - E_i)^2}{E_i}$
74	65.117	1.2118
20	28.883	2.7320
50	58.883	1.3491
35	26.117	3.0214

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 8.305$$

The number of degrees of freedom $\nu = (2-1)(2-1) = 1$; from the tables: $\chi_1^2(5\%) = 3.841$

As 8.305 is greater than 3.841, reject H_0 at the 5% level. There is evidence of an association between the age and number of credit cards possessed.

11 a H_0: There is no association between injury rate and choice of gym.
H_1: There is an association between injury rate and choice of gym.

b Calculate the expected value for each cell by multiplying column and row totals and dividing by the grand total:

$$E_{C,I} = \frac{175 \times 34}{865} = 6.88 \text{ (2 d.p.)}$$

c The test statistic (X^2) calculations are:

O_i	E_i	$\frac{(O_i - E_i)^2}{E_i}$
15	9.32	3.462
4	10.14	3.718
8	6.88	0.182
7	7.66	0.057
222	227.68	0.142
254	247.86	0.152
167	168.12	0.007
188	187.34	0.002

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 7.722$$

The number of degrees of freedom $\nu = (4-1)(2-1) = 3$; from the tables: $\chi_3^2(5\%) = 7.815$

As 7.722 is less than 7.815, there is insufficient evidence to reject H_0 at the 5% level. There is no reason to believe there is an association between injury rate and choice of gym.

12 a H_0: There is no association between science studied and salary.
H_1: There is an association between science studied and salary.

b Calculate the expected value for each cell:

$$E_{B,0-20} = \frac{104 \times 9}{323} = 2.90$$

The expected frequencies (E_i) are:

	Salary				
	£0–£20k	**£20k–£40k**	**£40–£60k**	**£60k–£80k**	**>£80k**
Biology	2.90	67.29	26.40	4.51	2.90
Chemistry	3.01	69.88	27.42	4.68	3.01
Physics	3.09	71.82	28.18	4.81	3.09

Require each cell of the expected table to have a value at least 5; merge the first two columns (so create a category £0–£40k) and the last two columns (for a category >£60k).

	Salary		
	£0–£40	**£40k–£60k**	**>£60k**
Biology	70.19	26.40	7.41
Chemistry	72.89	27.42	7.69
Physics	74.92	28.18	7.90

The test statistic (X^2) calculations are:

O_i	E_i	$\frac{(O_i - E_i)^2}{E_i}$
73	70.19	0.7985
23	26.40	0.0560
8	7.41	0.5486
75	72.89	0.7049
27	27.42	0.2033
6	7.69	2.5344
70	74.92	0.0419
32	28.18	1.0664
9	7.90	2.6112

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 2.030$$

The number of degrees of freedom $\nu = (3-1)(3-1) = 4$; from the tables: $\chi_4^2(5\%) = 9.488$

As 2.030 is less than 9.488, there is insufficient evidence to reject H_0 at the 5% level. There is no reason to believe there is an association between science studied and salary.