Contingency tables

Worked solutions

Further Maths topic practice

This topic online

1 A survey of 200 pupils recorded how they travel to school and their key stage.

	KS3	KS4	Total
Walk	45	35	80
Car	55	65	120
Total	100	100	200

A test is to be carried out to see whether there is an association between how a pupil travels and their key stage.

a. Find the expected frequency for each cell if there is no association.

b. State the number of degrees of freedom for the test.

(3 marks)

a) Totals: rows 80 and 120, columns 100 and 100

Expected frequency = row total × column total ÷ grand total

$$E(\text{Walk, KS3}) = \frac{80 \times 100}{200} = 40$$

$$E(\text{Walk, KS4}) = \frac{80 \times 100}{200} = 40$$

$$E(\text{Car, KS3}) = \frac{120 \times 100}{200} = 60$$

$$E(\text{Car, KS4}) = \frac{120 \times 100}{200} = 60$$

$\boxed{40, 40, 60, 60}$

b)

Degrees of freedom = (rows − 1)(columns − 1)

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

$\boxed{\nu = 1}$

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

2 Pupils at a school were asked to choose their favourite sport from four options.

	Football	Netball	Swimming	Tennis	Total
Year 7	22	15	18	10	65
Year 8	25	12	20	13	70
Year 9	18	20	14	13	65
Total	65	47	52	36	200

The PE department wants to test whether favourite sport is independent of year group.

a. Write down suitable hypotheses.

b. Find the expected frequency for Year 8 pupils who choose swimming.

c. State the degrees of freedom and the critical value for a test at the 5% level of significance.

(4 marks)

a)

H_0: favourite sport and year group are independent

H_1: favourite sport and year group are not independent

b) Year 8 row total and swimming column total

$$\text{Year 8 total} = 25 + 12 + 20 + 13 = 70$$

$$\text{Swimming total} = 18 + 20 + 14 = 52$$

$$E = \frac{70 \times 52}{200}$$

$$E = 18.20$$

18.20

c)

Degrees of freedom = (rows − 1)(columns − 1)

$$\nu = (3 - 1)(4 - 1) = 6$$

5% level, $\nu = 6$

critical value = 12.592

$\nu = 6$, critical value 12.592

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

3 A random sample of 150 households in a district was asked whether they own a pet. Each household was also classified by whether it is in a town or a village.

	Town	Village	Total
Has a pet	38	22	60
No pet	42	48	90
Total	80	70	150

Test, at the 5% level of significance, whether there is an association between owning a pet and living in a town or a village. State your hypotheses and show your working clearly. (7 marks)

H_0: owning a pet and living in a town or village are independent

H_1: there is an association between them

Expected frequency = row total × column total ÷ grand total

$$E(\text{Has a pet, Town}) = \frac{60 \times 80}{150} = 32$$

$$E(\text{Has a pet, Village}) = \frac{60 \times 70}{150} = 28$$

$$E(\text{No pet, Town}) = \frac{90 \times 80}{150} = 48$$

$$E(\text{No pet, Village}) = \frac{90 \times 70}{150} = 42$$

Each contribution = $(O - E)^2 \div E$

$$\text{Has a pet, Town: } \frac{(38 - 32)^2}{32} = 1.125$$

$$\text{Has a pet, Village: } \frac{(22 - 28)^2}{28} = 1.286$$

$$\text{No pet, Town: } \frac{(42 - 48)^2}{48} = 0.750$$

$$\text{No pet, Village: } \frac{(48 - 42)^2}{42} = 0.857$$

$$\chi^2 = 1.125 + 1.286 + 0.750 + 0.857$$

$$\chi^2 = 4.018$$

Degrees of freedom = (rows − 1)(columns − 1)

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

Critical value, 5% level, $\nu = 1$

critical value = 3.841

$4.018 > 3.841$, so reject H_0

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

Evidence of an association between owning a pet and where a household lives

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

4 A travel agent asked 240 customers where they went on their main holiday last year.

Age group	UK	Europe	Further afield	Total
Under 40	30	52	38	120
40 and over	50	38	32	120
Total	80	90	70	240

Some of the expected frequencies for a test of independence are shown below.

Expected	UK	Europe	Further afield
Under 40	40	45	
40 and over		45	

a. Complete the table of expected frequencies.

b. Test, at the 5% level of significance, whether there is an association between age group and holiday destination. State your hypotheses and the critical value clearly.

(8 marks)

a) Row totals are both 120

$$E(\text{Under 40, Further afield}) = \frac{120 \times 70}{240} = 35$$

$$E(\text{40 and over, UK}) = \frac{120 \times 80}{240} = 40$$

$$E(\text{40 and over, Further afield}) = \frac{120 \times 70}{240} = 35$$

40, 45, 35 in each row

b)

H_0: age group and holiday destination are independent

H_1: they are not independent

Each contribution = $(O - E)^2 \div E$

$$\text{Under 40, UK: } \frac{(30 - 40)^2}{40} = 2.500$$

$$\text{Under 40, Europe: } \frac{(52 - 45)^2}{45} = 1.089$$

$$\text{Under 40, Further afield: } \frac{(38 - 35)^2}{35} = 0.257$$

$$\text{40 and over, UK: } \frac{(50 - 40)^2}{40} = 2.500$$

$$\text{40 and over, Europe: } \frac{(38 - 45)^2}{45} = 1.089$$

$$\text{40 and over, Further afield: } \frac{(32 - 35)^2}{35} = 0.257$$

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

$$\chi^2 = 2.500 + 1.089 + 0.257 + 2.500 + 1.089 + 0.257$$

$$\chi^2 = 7.692$$

Degrees of freedom = (rows − 1)(columns − 1)

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

Critical value, 5% level, $\nu = 2$

critical value = 5.991

$7.692 > 5.991$, so reject H_0

Evidence that holiday choice is associated with age group

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

5 A school nurse asks pupils how long they usually sleep on a school night.

	Less than 7 h	7 to 8 h	More than 8 h	Total
Year 9	12	30	28	70
Year 11	20	34	16	70
Year 13	25	22	8	55
Total	57	86	52	195

She wants to test whether sleep time is independent of year group.

a. Without calculating every expected frequency, identify the cell with the smallest expected frequency, giving a reason, and calculate it.

b. Explain what this tells the nurse about combining classes.

c. The test statistic is 17.09. Test, at the 1% level of significance, whether sleep time is independent of year group. State your hypotheses and critical value.

(6 marks)

a) Row and column totals

Rows: 70, 70, 55

Columns: 57, 86, 52

Smallest E = smallest row total × smallest column total ÷ N

$$E(\text{Year 13, More than 8 h}) = \frac{55 \times 52}{195}$$

$$E = 14.67$$

Year 13, more than 8 h: 14.67

b)

14.67 ≥ 5, so every expected frequency is at least 5

No classes need combining

c)

H_0: sleep time is independent of year group

H_1: sleep time is not independent of year group

Degrees of freedom = (rows − 1)(columns − 1)

$$\nu = (3-1)(3-1) = 4$$

1% level, $\nu = 4$

critical value = 13.277

17.09 > 13.277, so reject H_0

Sleep time is associated with year group

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

6 Pupils at three secondary schools in a city were asked how they usually travel to school.

	Walk	Bus	Car	Cycle	Total
School P	40	25	30	5	100
School Q	28	40	22	2	92
School R	32	35	38	3	108
Total	100	100	90	10	300

...

(10 marks)

a) Cycle column total = 5 + 2 + 3 = 10, grand total 300

$$E(\text{School P, Cycle}) = \frac{100 \times 10}{300} = 3.33$$

$$E(\text{School Q, Cycle}) = \frac{92 \times 10}{300} = 3.07$$

$$E(\text{School R, Cycle}) = \frac{108 \times 10}{300} = 3.60$$

All less than 5

Combine Cycle with Walk, as the question says

Combine Cycle with Walk: both are active travel

b) New observed table: Walk or cycle = Walk + Cycle

School P: 45, 25, 30

School Q: 30, 40, 22

School R: 35, 35, 38

H_0: method of travel is independent of school

H_1: method of travel is not independent of school

Expected frequency = row total × column total ÷ grand total

$$E(\text{School P, Walk or cycle}) = \frac{100 \times 110}{300} = 36.67$$

$$E(\text{School P, Bus}) = \frac{100 \times 100}{300} = 33.33$$

$$E(\text{School P, Car}) = \frac{100 \times 90}{300} = 30$$

$$E(\text{School Q, Walk or cycle}) = \frac{92 \times 110}{300} = 33.73$$

$$E(\text{School Q, Bus}) = \frac{92 \times 100}{300} = 30.67$$

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

$$E(\text{School Q, Car}) = \frac{92 \times 90}{300} = 27.60$$

$$E(\text{School R, Walk or cycle}) = \frac{108 \times 110}{300} = 39.60$$

$$E(\text{School R, Bus}) = \frac{108 \times 100}{300} = 36$$

$$E(\text{School R, Car}) = \frac{108 \times 90}{300} = 32.40$$

Each contribution = $(O - E)^2 \div E$

$$\text{School P, Walk or cycle: } \frac{(45 - 36.67)^2}{36.67} = 1.894$$

$$\text{School P, Bus: } \frac{(25 - 33.33)^2}{33.33} = 2.083$$

$$\text{School P, Car: } \frac{(30 - 30)^2}{30} = 0.000$$

$$\text{School Q, Walk or cycle: } \frac{(30 - 33.73)^2}{33.73} = 0.413$$

$$\text{School Q, Bus: } \frac{(40 - 30.67)^2}{30.67} = 2.841$$

$$\text{School Q, Car: } \frac{(22 - 27.60)^2}{27.60} = 1.136$$

$$\text{School R, Walk or cycle: } \frac{(35 - 39.60)^2}{39.60} = 0.534$$

$$\text{School R, Bus: } \frac{(35 - 36)^2}{36} = 0.028$$

$$\text{School R, Car: } \frac{(38 - 32.40)^2}{32.40} = 0.968$$

$$\chi^2 = 1.894 + 2.083 + 0.000 + 0.413 + 2.841 + 1.136 + 0.534 + 0.028 + 0.968$$

$$\chi^2 = 9.897$$

Degrees of freedom = (rows - 1)(columns - 1)

$$\nu = (3 - 1)(3 - 1) = 4$$

Critical value, 5% level, $\nu = 4$

critical value = 9.488

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

$9.897 > 9.488$, so reject H_0

Evidence that method of travel is associated with school

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

7 A café owner records whether each of a random sample of 250 customers bought a hot drink or a cold drink, and the time of day.

	Morning	Afternoon	Evening	Total
Hot drink	64	40	26	130
Cold drink	36	50	34	120
Total	100	90	60	250

The contributions $(O - E)^2/E$ to the test statistic are shown below, correct to 3 decimal places.

	Morning	Afternoon	Evening
Hot drink	2.769	0.988	0.867
Cold drink	c	1.070	0.939

a. Show that the expected frequency for cold drinks in the morning is 48 and find the value of c.

b. Test, at the 5% level of significance, whether type of drink is independent of time of day. State your hypotheses and the critical value.

c. Use the contributions to describe the association.

(8 marks)

a) Cold drink row total and morning column total

$$E = \frac{120 \times 100}{250}$$

$$E = 48$$

$$c = \frac{(36 - 48)^2}{48}$$

$$c = \frac{144}{48} = 3.000$$

$$\boxed{c = 3.000}$$

b)

H_0: type of drink is independent of time of day

H_1: type of drink is not independent of time of day

Add all six contributions

$$\chi^2 = 2.769 + 0.988 + 0.867 + 3.000 + 1.070 + 0.939$$

$$\chi^2 = 9.633$$

Degrees of freedom = (rows - 1)(columns - 1)

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

Critical value, 5% level, $\nu = 2$

critical value = 5.991

$9.633 > 5.991$, so reject H_0

$$\boxed{\text{Evidence that drink type is associated with time of day}}$$

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

c) Largest contributions are in the morning column

Hot drink, morning: O = 64, E = 52 (more than expected)

Cold drink, morning: O = 36, E = 48 (fewer than expected)

Hot drinks are more popular in the morning

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

8 After a local council election, a researcher asked a random sample of 390 adults whether they had voted.

Age group	18 to 29	30 to 44	45 to 64	65 and over	Total
Voted	30	48	70	62	210
Did not vote	50	52	50	28	180
Total	80	100	120	90	390

Test, at the 1% level of significance, whether there is an association between age group and voting. State your hypotheses and the critical value clearly, and show your working. (9 marks)

H_0: voting and age group are independent

H_1: voting and age group are not independent

Expected frequency = row total × column total ÷ grand total

$$E(\text{Voted, 18 to 29}) = \frac{210 \times 80}{390} = 43.08$$

$$E(\text{Voted, 30 to 44}) = \frac{210 \times 100}{390} = 53.85$$

$$E(\text{Voted, 45 to 64}) = \frac{210 \times 120}{390} = 64.62$$

$$E(\text{Voted, 65 and over}) = \frac{210 \times 90}{390} = 48.46$$

$$E(\text{Did not vote, 18 to 29}) = \frac{180 \times 80}{390} = 36.92$$

$$E(\text{Did not vote, 30 to 44}) = \frac{180 \times 100}{390} = 46.15$$

$$E(\text{Did not vote, 45 to 64}) = \frac{180 \times 120}{390} = 55.38$$

$$E(\text{Did not vote, 65 and over}) = \frac{180 \times 90}{390} = 41.54$$

Each contribution = $(O - E)^2 \div E$

$$\text{Voted, 18 to 29: } \frac{(30 - 43.08)^2}{43.08} = 3.970$$

$$\text{Voted, 30 to 44: } \frac{(48 - 53.85)^2}{53.85} = 0.635$$

$$\text{Voted, 45 to 64: } \frac{(70 - 64.62)^2}{64.62} = 0.449$$

$$\text{Voted, 65 and over: } \frac{(62 - 48.46)^2}{48.46} = 3.782$$

$$\text{Did not vote, 18 to 29: } \frac{(50 - 36.92)^2}{36.92} = 4.631$$

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

Did not vote, 30 to 44: $\frac{(52 - 46.15)^2}{46.15} = 0.741$

Did not vote, 45 to 64: $\frac{(50 - 55.38)^2}{55.38} = 0.524$

Did not vote, 65 and over: $\frac{(28 - 41.54)^2}{41.54} = 4.413$

$$\chi^2 = 3.970 + 0.635 + 0.449 + 3.782 + 4.631 + 0.741 + 0.524 + 4.413$$

$$\chi^2 = 19.143$$

Degrees of freedom = (rows - 1)(columns - 1)

$\nu = (2 - 1)(4 - 1) = 3$

Critical value, 1% level, $\nu = 3$

critical value = 11.345

$19.143 > 11.345$, so reject H_0

Evidence at the 1% level that voting is associated with age group

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

9 A sixth form college runs a pilot survey asking 40 students whether they prefer online or face to face revision sessions.

Pilot	Year 12	Year 13	Total
Online	14	6	20
Face to face	10	10	20
Total	24	16	40

a. Test, at the 5% level of significance, whether preference is independent of year group.

The college then surveys 120 students and, by chance, every frequency is exactly three times the pilot frequency.

b. Show that the test statistic for the larger survey is three times the test statistic for the pilot, and complete the test at the 5% level.

c. Explain why the test should not be carried out on the percentages of students in each cell. (10 marks)

a)

H_0: preference is independent of year group

H_1: preference is not independent of year group

Expected frequency = row total × column total ÷ grand total

$$E(\text{Online, Year 12}) = \frac{20 \times 24}{40} = 12$$

$$E(\text{Online, Year 13}) = \frac{20 \times 16}{40} = 8$$

$$E(\text{Face to face, Year 12}) = \frac{20 \times 24}{40} = 12$$

$$E(\text{Face to face, Year 13}) = \frac{20 \times 16}{40} = 8$$

Each contribution = $(O - E)^2 \div E$

$$\text{Online, Year 12: } \frac{(14 - 12)^2}{12} = 0.333$$

$$\text{Online, Year 13: } \frac{(6 - 8)^2}{8} = 0.500$$

$$\text{Face to face, Year 12: } \frac{(10 - 12)^2}{12} = 0.333$$

$$\text{Face to face, Year 13: } \frac{(10 - 8)^2}{8} = 0.500$$

$$\chi^2 = 0.333 + 0.500 + 0.333 + 0.500$$

$$\chi^2 = 1.667$$

Degrees of freedom = (rows − 1)(columns − 1)

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

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

Critical value, 5% level, $\nu = 1$

critical value $= 3.841$

$1.667 < 3.841$, so do not reject H_0

No evidence of an association in the pilot survey

b) Every O x 3, so every total x 3

$$\text{new } E = \frac{3R \times 3C}{3N} = 3 \times \frac{R \times C}{N} = 3E$$

$$\text{new contribution} = \frac{(3O - 3E)^2}{3E}$$

$$\text{new contribution} = \frac{9(O-E)^2}{3E} = 3 \times \frac{(O-E)^2}{E}$$

$$\chi^2 = 3 \times \frac{5}{3} = 5.000$$

$5.000 > 3.841$, so reject H_0

Evidence of an association in the larger survey

c) χ^2 changes with the size of the frequencies

Must use frequencies: percentages give a different χ^2

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

10 A library service asked a random sample of 270 adults and children how often they visit a library.

Age	Weekly	Monthly	Rarely	Total
Under 18	4	3	5	12
18 to 34	20	26	34	80
35 to 59	30	32	28	90
60 and over	46	24	18	88
Total	100	85	85	270

a. Explain why the Under 18 row must be combined with another row before a test can be carried out, and why combining it with the 18 to 34 row is sensible.

b. Test, at the 5% level of significance, whether frequency of library visits is independent of age. State your hypotheses, the degrees of freedom and the critical value.

c. Describe the association, using the cells that contribute most.

(10 marks)

a) Under 18 row total = 12

$$E(\text{Under 18, Weekly}) = \frac{12 \times 100}{270} = 4.44$$

$$E(\text{Under 18, Monthly}) = \frac{12 \times 85}{270} = 3.78$$

$$E(\text{Under 18, Rarely}) = \frac{12 \times 85}{270} = 3.78$$

All less than 5, so combine the row

Join with 18 to 34: the next age group

b) New table

Under 35: 24, 29, 39

35 to 59: 30, 32, 28

60 and over: 46, 24, 18

H_0: frequency of visits is independent of age

H_1: frequency of visits is not independent of age

Expected frequency = row total × column total ÷ grand total

$$E(\text{Under 35, Weekly}) = \frac{92 \times 100}{270} = 34.07$$

$$E(\text{Under 35, Monthly}) = \frac{92 \times 85}{270} = 28.96$$

$$E(\text{Under 35, Rarely}) = \frac{92 \times 85}{270} = 28.96$$

$$E(\text{35 to 59, Weekly}) = \frac{90 \times 100}{270} = 33.33$$

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

$$E(35 \text{ to } 59, \text{Monthly}) = \frac{90 \times 85}{270} = 28.33$$

$$E(35 \text{ to } 59, \text{Rarely}) = \frac{90 \times 85}{270} = 28.33$$

$$E(60 \text{ and over}, \text{Weekly}) = \frac{88 \times 100}{270} = 32.59$$

$$E(60 \text{ and over}, \text{Monthly}) = \frac{88 \times 85}{270} = 27.70$$

$$E(60 \text{ and over}, \text{Rarely}) = \frac{88 \times 85}{270} = 27.70$$

Each contribution = $(O - E)^2 \div E$

$$\text{Under 35, Weekly: } \frac{(24 - 34.07)^2}{34.07} = 2.978$$

$$\text{Under 35, Monthly: } \frac{(29 - 28.96)^2}{28.96} = 0.000$$

$$\text{Under 35, Rarely: } \frac{(39 - 28.96)^2}{28.96} = 3.478$$

$$\text{35 to 59, Weekly: } \frac{(30 - 33.33)^2}{33.33} = 0.333$$

$$\text{35 to 59, Monthly: } \frac{(32 - 28.33)^2}{28.33} = 0.475$$

$$\text{35 to 59, Rarely: } \frac{(28 - 28.33)^2}{28.33} = 0.004$$

$$\text{60 and over, Weekly: } \frac{(46 - 32.59)^2}{32.59} = 5.515$$

$$\text{60 and over, Monthly: } \frac{(24 - 27.70)^2}{27.70} = 0.495$$

$$\text{60 and over, Rarely: } \frac{(18 - 27.70)^2}{27.70} = 3.399$$

$$\chi^2 = 2.978 + 0.000 + 3.478 + 0.333 + 0.475 + 0.004 + 5.515 + 0.495 + 3.399$$

$$\chi^2 = 16.678$$

Degrees of freedom = (rows - 1)(columns - 1)

$$\nu = (3 - 1)(3 - 1) = 4$$

Critical value, 5% level, $\nu = 4$

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

critical value = 9.488

16.678 > 9.488, so reject H_0

Evidence that how often people use the library is associated with age

c) Largest contributions

60 and over, Weekly: 5.515 (O = 46 > E = 32.59)

Under 35, Rarely: 3.478 (O = 39 > E = 28.96)

Older people visit more often; under 35s visit rarely

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

11 A survey of 200 pupils recorded whether they usually have a school dinner or a packed lunch.

	KS3	KS4	Total
School dinner	x	50	
Packed lunch	$90 - x$	60	
Total	90	110	200

For a test of independence, the expected frequency for KS3 pupils having a school dinner is 45.

a. Find the value of x.

b. Test, at the 5% level of significance, whether type of lunch is independent of key stage. State your hypotheses and the critical value.

(8 marks)

a) School dinner row total = $x + 50$

$$\frac{(x + 50) \times 90}{200} = 45$$

$$(x + 50) \times 90 = 9000$$

$$x + 50 = 100$$

$$x = 50$$

$$\boxed{x = 50}$$

b) Observed table: 50, 50 / 40, 60

H_0: type of lunch is independent of key stage

H_1: type of lunch is not independent of key stage

Expected frequency = row total × column total ÷ grand total

$$E(\text{School dinner, KS3}) = \frac{100 \times 90}{200} = 45$$

$$E(\text{School dinner, KS4}) = \frac{100 \times 110}{200} = 55$$

$$E(\text{Packed lunch, KS3}) = \frac{100 \times 90}{200} = 45$$

$$E(\text{Packed lunch, KS4}) = \frac{100 \times 110}{200} = 55$$

Each contribution = $(O - E)^2 \div E$

$$\text{School dinner, KS3: } \frac{(50 - 45)^2}{45} = 0.556$$

$$\text{School dinner, KS4: } \frac{(50 - 55)^2}{55} = 0.455$$

$$\text{Packed lunch, KS3: } \frac{(40 - 45)^2}{45} = 0.556$$

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

Packed lunch, KS4: $\frac{(60-55)^2}{55} = 0.455$

$\chi^2 = 0.556 + 0.455 + 0.556 + 0.455$

$\chi^2 = 2.020$

Degrees of freedom = (rows − 1)(columns − 1)

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

Critical value, 5% level, $\nu = 1$

critical value = 3.841

$2.020 < 3.841$, so do not reject H_0

No evidence that lunch type is associated with key stage

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

12 A head of maths records, for 190 Year 11 pupils, whether they used practice papers in their revision and the grade they achieved in GCSE Mathematics.

	Grade 4 or below	Grade 5 or 6	Grade 7 or above	Total
Used practice papers	18	42	40	100
Did not	32	38	20	90
Total	50	80	60	190

a. Test, at the 5% level of significance, whether grade achieved is independent of using practice papers. State your hypotheses and the critical value clearly.

b. Identify the two cells with the largest contributions and say what they show.

c. The head of maths says this proves that using practice papers causes higher grades. Comment on this claim.

(10 marks)

a)

H_0: grade achieved is independent of using practice papers

H_1: grade achieved is not independent of using practice papers

Expected frequency = row total × column total ÷ grand total

$$E(\text{Used practice papers, Grade 4 or below}) = \frac{100 \times 50}{190} = 26.32$$

$$E(\text{Used practice papers, Grade 5 or 6}) = \frac{100 \times 80}{190} = 42.11$$

$$E(\text{Used practice papers, Grade 7 or above}) = \frac{100 \times 60}{190} = 31.58$$

$$E(\text{Did not, Grade 4 or below}) = \frac{90 \times 50}{190} = 23.68$$

$$E(\text{Did not, Grade 5 or 6}) = \frac{90 \times 80}{190} = 37.89$$

$$E(\text{Did not, Grade 7 or above}) = \frac{90 \times 60}{190} = 28.42$$

Each contribution = $(O - E)^2 \div E$

$$\text{Used practice papers, Grade 4 or below: } \frac{(18 - 26.32)^2}{26.32} = 2.628$$

$$\text{Used practice papers, Grade 5 or 6: } \frac{(42 - 42.11)^2}{42.11} = 0.000$$

$$\text{Used practice papers, Grade 7 or above: } \frac{(40 - 31.58)^2}{31.58} = 2.246$$

$$\text{Did not, Grade 4 or below: } \frac{(32 - 23.68)^2}{23.68} = 2.920$$

$$\text{Did not, Grade 5 or 6: } \frac{(38 - 37.89)^2}{37.89} = 0.000$$

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

Did not, Grade 7 or above: $\frac{(20 - 28.42)^2}{28.42} = 2.495$

$\chi^2 = 2.628$

$+$

0.000

$+$

2.246

$+$

2.920

$+$

0.000

$+$

2.495

$\chi^2 = 10.289$

Degrees of freedom = (rows - 1)(columns - 1)

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

Critical value, 5% level, $\nu = 2$

critical value = 5.991

$10.289 > 5.991$, so reject H_0

Evidence that grade is associated with using practice papers

b) Largest contributions

Did not, grade 4 or below: 2.920 ($O = 32 > E = 23.68$)

Used, grade 4 or below: 2.628 ($O = 18 < E = 26.32$)

Non-users got grade 4 or below more often than expected

c) What a χ^2 test can show

It shows an association, not a cause

Not proof: other factors (e.g. motivation) may cause both

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.