inkMaths

ANSWER BOOK

Contingency tables

Further Statistics 1 · Year FM

7 questions · 23 marks

NAVY EDITION

© 2026 InkMaths. All rights reserved.
inkphysics.co.uk/maths/ · youtube.com/@InkPhysics

SECTION A · Warm-up

A1 H_0 is that the two variables are independent, so there is no association between them. Each expected frequency is row total × column total ÷ grand total. B1 for the hypothesis, B1 for the formula. [2]

A2 $(4 - 1)(5 - 1) =$ **12**. B1. No further subtraction is made for estimated parameters. The row and column totals have already absorbed them, which is the difference between this test and a goodness-of-fit test. [1]

SECTION B · Standard

B1 Row totals are 50 and 50, column totals 40, 40 and 20, grand total 100. Each expected frequency is row total × column total ÷ 100, giving 50 × 40/100 = **20** for the first two columns of each row and 50 × 20/100 = **10** for the third. M1 for the row and column totals, M1 for row × column ÷ grand total, A1 for all six values. Equal row totals make both rows expect the same split. Set the six values out in a second table alongside the observed one, since that table is where the method mark sits. [3]

B2 H_0: no association between the two variables; H_1: the variables are associated. The contributions $(O - E)^2/E$ are 25/20, 25/20, 0, 25/20, 25/20 and 0, totalling **5.0**. There are $(2 - 1)(3 - 1) =$ **2** degrees of freedom, and $5.0 < 5.991$, so the statistic is not in the critical region. **Do** not reject H_0. There is insufficient evidence at the 5% level of an association, although the value sits fairly close to the boundary. B1 for the hypotheses, M1 for the contributions, A1 for 5.0, B1 for 2 degrees of freedom, A1 for the comparison and conclusion in context. Quote every contribution. A bare total loses the working marks. [5]

B3 Once both row totals and both column totals are fixed, choosing any single cell determines the other three by subtraction. Exactly one cell is free to vary, and $(2 - 1)(2 - 1) = 1$ records that. B1 for the totals fixing the remaining cells, B1 for exactly one free cell. [2]

B4 It may be concluded that there is evidence of an association between the two variables, since the observed pattern is unlikely if they were independent. It may not be concluded that one variable causes the other, nor that the association is large. B1 for evidence of association, B1 for ruling out a causal conclusion, B1 for the strength not being measured. The test measures evidence against independence, not the strength or direction of any effect, and a large enough sample makes even a slight association significant. [3]

© 2026 InkMaths. All rights reserved.

SECTION C · Stretch

C1 Row totals 75 and 45, column totals 60, 50 and 10, grand total 120. Expected frequencies: row 1 gives **37.5,** 31.25, 6.25; row 2 gives **22.5,** 18.75, 3.75. [7]

The last cell of row 2 has an expected frequency of 3.75, below 5, so the third column must be **merged** with the second. The observed table becomes (40, 35) and (20, 25), with expected frequencies 37.5, 37.5, 22.5 and 22.5, all above 5.

Merging changes the size of the table, so the degrees of freedom must be recounted from the merged table: $(2 - 1)(2 - 1) = \mathbf{1}$, not 2.

$\chi^2 = 2 \times 6.25/37.5 + 2 \times 6.25/22.5 = 0.333 + 0.556 = \mathbf{0.889}$. Since $0.889 < 3.841$, **do** not reject H_0. There is insufficient evidence at the 5% level of an association. M1 for row × column ÷ grand total, A1 for all six expected frequencies, B1 for spotting the expected frequency below 5, M1 for merging the last two columns, B1 for 1 degree of freedom, M1 for the χ^2 calculation, A1 for 0.889 and the conclusion. Using 5.991 with the merged table is the standard error here. The degrees of freedom follow the table you actually test, not the one you were given.

© 2026 InkMaths. All rights reserved.