inkMaths

ANSWER BOOK

# Goodness-of-fit tests

Further Statistics 1 · Year FM

6 questions · 20 marks

NAVY EDITION

## SECTION A · Warm-up

**A1** [2] $\chi^2 = \Sigma(O - E)^2/E$. Dividing by E scales each gap against the size of the class it sits in, so a discrepancy of 3 where 5 were expected counts far more heavily than the same gap where 500 were expected.

**A2** [3] Each expected frequency is 100/5 = 20. The squared differences are 25, 4, 4, 25, 0, totalling 58. Dividing by 20 gives $\chi^2 = 2.9$.

## SECTION B · Standard

**B1** [3] Five classes with no parameter estimated: 5 − 1 = 4 degrees of freedom. Since 2.9 < 9.488, do not reject $H_0$: there is insufficient evidence at the 5% level that the spinner is biased. That is not proof of fairness, only an absence of evidence against it.

**B2** [4] The estimate is the sample mean: $(0 \times 12 + 1 \times 18 + 2 \times 15 + 3 \times 9 + 4 \times 6)/60 = 99/60 = 1.65$. Five classes, minus one for the total, minus one more for the estimated $\lambda$: 3 degrees of freedom.

**B3** [4] Any class with an expected frequency below 5 is combined with a neighbour, repeatedly if necessary, until every expected frequency reaches 5. Small expected values inflate the $(O - E)^2/E$ terms and distort the statistic. Pooling reduces the number of classes, and the degrees of freedom are counted from the pooled classes, not the original ones.

## SECTION C · Stretch

**C1** [4] Specifying the parameter in advance gives one more degree of freedom, since nothing was estimated. Estimating it from the data lets the model bend towards the observations, so the expected frequencies sit closer to what was seen and the squared gaps shrink. Surrendering a degree of freedom is the price of that flexibility, and it keeps the test honest.