Past-paper walk-through

Chi-squared tests ▪ 4.3

eXercise

Questions in this set

- 9231/41 N25 Q2 ▪ 6 marks
- 9231/42 N25 Q3 ▪ 8 marks
- 9231/44 N25 Q5 ▪ 7 marks
- 9231/41 J25 Q3 ▪ 8 marks
- 9231/43 J25 Q1 ▪ 6 marks

Reading the mark scheme

- **B1** a mark that stands on its own – the point is either made or not.
- **M1** a *method* mark: the right approach, even if the number is wrong.
- **A1** an *answer* mark, and it usually needs its M mark first.
- **C1** a working mark on the way to the answer.
- // separates answers the examiner accepts equally.
- **ecf** = error carried forward: a later part is marked on your own earlier answer.
- **owtte** = "or words to that effect" – the wording need not match.

Question 2

9231/41 N25 ▪ Q2 ▪ 6 marks

The manager of a car park claims that the number of cars entering the car park follows a Poisson distribution with mean 2.8. The numbers of cars entering the car park are recorded on a working day during successive 5-minute periods. The table contains the observed frequencies, together with most of the expected frequencies and their contributions to the χ^2-test statistic (with p, q the missing expected frequency and contribution for 3 cars).

(a) Find the value of p and the value of q. **[2]**

(b) Carry out a goodness of fit test at the 5% significance level to investigate the manager's claim. **[4]**

Question 2

Number of cars	0	1	2	3	4	5	$\geqslant 6$
Observed frequency	2	15	31	29	13	3	7
Expected frequency	6.081	17.03	23.84	p	15.57	8.721	6.511
χ^2-test statistic	2.739	0.241	2.152	q	0.425	3.753	0.037

Solution 2

(a)

Mark scheme

- $p = \dfrac{e^{-2.8}(2.8)^3}{3!} \times 100 = 22.248$ (AWRT 22.25). $q = \dfrac{(29 - 22.25)^2}{22.25} = 2.05$.

Solution 2

(b)

Mark scheme

- H_0: Po(2.8) fits the data; H_1: Po(2.8) does not fit the data. $\chi^2 = 2.739 + 0.241 + 2.152 + 2.049 + 0.425 + 3.753 + 0.037 = 11.4$. With 6 degrees of freedom, the critical value is $\chi^2_{0.95}(6) = 12.59$. Since $11.4 < 12.59$, do not reject H_0: there is insufficient evidence to suggest that Po(2.8) is not a good fit (i.e. that the manager's claim is false).

Question 3

9231/42 N25 ▪ Q3 ▪ 8 marks

A traffic expert claims that the number of breakdowns occurring each day on a busy section of motorway follows a Poisson distribution with mean 0.7. Over a 200-day period the observed frequencies for 0, 1, 2, 3, 4, $\geqslant 5$ breakdowns were 88, 73, 26, 7, 3, 3, with expected frequencies 99.317, *m*, 24.333, 5.678, 0.994, *n*.

(a) Find the value of *m* and the value of *n*, correct to 3 decimal places. **[2]**

(b) Carry out a goodness of fit test at the 5% significance level to investigate the expert's claim. **[6]**

Question 3

Number of breakdowns per day	0	1	2	3	4	$\geqslant 5$
Observed frequency	88	73	26	7	3	3
Expected frequency	99.317	m	24.333	5.678	0.994	n

Solution 3

(a)

Mark scheme

- $m = 200\, e^{-0.7}(0.7) = 69.522$;
  $n = 200 - (99.317 + 69.522 + 24.333 + 5.678 + 0.994) = 0.156$.

Solution 3

(b)

Mark scheme

- Combine cells so each expected $\geqslant 5$: classes 0, 1, 2, $\geqslant 3$ with observed 88, 73, 26, 13 and expected 99.317, 69.522, 24.333, 6.828. $\chi^2 = 1.290 + 0.174 + 0.114 + 5.579 = 7.16$. With $\nu = 3$, $\chi^2_{0.05} = 7.815$; since $7.16 < 7.815$, do not reject H_0 — the Poisson(0.7) model fits.

Question 5

9231/44 N25 ▪ Q5 ▪ 7 marks

A driving instructor believes that the performance (pass or fail) of students when taking a driving test is associated with their age. The following table summarises the number of students who pass and who fail, and the ages in years of the students taking the test, over a period of three years. Pass: under 20 $= 34$, 20–40 $= 41$, over 40 $= 6$ (total 81). Fail: under 20 $= 16$, 20–40 $= 39$, over 40 $= 9$ (total 64). Column totals: 50, 80, 15 (grand total 145). Test, at the 10% significance level, whether performance is independent of age of student. **[7]**

Question 5

	age of student			
	under 20	20–40	over 40	total
pass	34	41	6	81
fail	16	39	9	64
total	50	80	15	145

Solution 5

Mark scheme

- H_0: performance is independent of age; H_1: performance is not independent of age. Expected frequencies: pass $27.93, 44.69, 8.38$; fail $22.07, 35.31, 6.62$. $\chi^2 = \dfrac{(34-27.93)^2}{27.93} + \cdots + \dfrac{(9-6.62)^2}{6.62} = 5.21$. With $(2-1)(3-1) = 2$ degrees of freedom, the critical value is $\chi^2_{0.90}(2) = 4.605$. Since $5.21 > 4.605$, reject H_0: there is sufficient evidence that performance and age are not independent (i.e. they are associated).

Question 3

9231/41 J25 ▪ Q3 ▪ 8 marks

Eggs in a supermarket are sold in boxes of six. A supermarket manager wishes to model the number of broken eggs in the boxes sold in the store. A random sample of 2000 boxes is taken and the number of broken eggs recorded. Observed frequencies (0–6 broken eggs): 1844, 143, 11, 0, 1, 0, 1.

(a) Use the data to estimate the probability that an egg is broken. Give your answer correct to 4 significant figures. **[1]**

(b) It is decided to carry out a goodness of fit test at the 0.5% significance level to determine whether a binomial distribution fits the data. The expected frequency for 1 broken egg is a. Show that $a = 162.6$ correct to 1 decimal place. **[1]**

(c) Carry out a goodness of fit test at the 0.5% level of significance to test whether a binomial distribution is a satisfactory model for the data. **[5]**

Question 3

(d) Give a reason why a binomial distribution may not be a suitable model in this situation. **[1]**

Question 3

Number of broken eggs	0	1	2	3	4	5	6
Observed frequency	1844	143	11	0	1	0	1
Expected frequency	1831.3	a	6.016	0.119	0.001	0.000	0.000

Question 3

Number of broken eggs	0	1	2	3	4	5	6
Observed frequency	1844	143	11	0	1	0	1

Solution 3

(a)

Mark scheme

- $\mathrm{P}(\text{egg broken}) = \dfrac{0 \cdot 1844 + 1 \cdot 143 + 2 \cdot 11 + 3 \cdot 0 + 4 \cdot 1 + 5 \cdot 0 + 6 \cdot 1}{2000 \times 6} = \dfrac{175}{12000} = \frac{7}{480} \approx 0.01458.$

Solution 3

(b)

Mark scheme

- With $X \sim \mathrm{B}\left(6, \frac{7}{480}\right)$, the expected frequency for 1 broken egg is $a = 2000 - (1831.3 + 6.016 + 0.119 + 0.001) = 162.6$ (AG). Equivalently $a = \binom{6}{1} \cdot \frac{7}{480}\left(1 - \frac{7}{480}\right)^5 \times 2000 = 162.6.$

Solution 3

(c) Mark scheme

- H_0: a binomial distribution is a satisfactory model; H_1: it is not. Combining the tail cells ($\geqslant 2$ broken: observed 13, expected 6.136), $\chi^2 = \dfrac{(1844 - 1831.3)^2}{1831.3} + \dfrac{(143 - 162.6)^2}{162.6} + \dfrac{(13 - 6.136)^2}{6.136} =$ $0.0881 + 2.363 + 7.678 = 10.1$. With 3 cells and 1 estimated parameter, $\nu = 3 - 1 - 1 = 1$, critical $\chi^2_{0.995}(1) = 7.879$. Since $10.1 > 7.879$, reject H_0: there is sufficient evidence that the manager's (binomial) model is not satisfactory.

Solution 3

(d)

Mark scheme

- The trials may not be independent, or the probability of an egg being broken may not be constant (e.g. boxes that are dropped have several broken eggs).

Question 1

9231/43 J25 ▪ Q1 ▪ 6 marks

A person's eye colour may be categorised as `brown'',` blue" or "other". A scientist claims that these eye colours are uniformly distributed and hence are equally likely to occur in the population. A survey of 120 people from this population found that 38 people had brown eyes, 52 people had blue eyes and 30 people had eyes which were neither brown nor blue. Use the data to carry out a goodness of fit test at the 5% significance level to test the scientist's claim. **[6]**

Solution 1

Mark scheme

- H_0: a uniform distribution fits the data; H_1: a uniform distribution does not fit the data. Expected frequencies are $40, 40, 40$.
$\chi^2 = \dfrac{(38-40)^2}{40} + \dfrac{(52-40)^2}{40} + \dfrac{(30-40)^2}{40} = 0.1 + 3.6 + 2.5 = 6.2$. With $3 - 1 = 2$ degrees of freedom, the critical value is $\chi^2_{0.95}(2) = 5.991$. Since $6.2 > 5.991$, reject H_0: there is sufficient evidence that the scientist's claim is incorrect (i.e. eye colour is not uniformly distributed).