Topic: Chi-Squared Tests (1)

Chapter Reference: Further Statistics 1, Chapter 6

7 minutes

1. A quality control manager regularly samples 20 items from a production line and records the number of defective items x. The results of 100 such samples are given in Table 1 below.

x	0	1	2	3	4	5	6	7 or more
Frequency	17	31	19	14	9	7	3	0

Table 1

The manager claimed that the number of defective items in a sample of 20 can be modelled by a binomial distribution. He used $X \sim B(20, 0.1)$ to calculate the expected frequencies given in Table 2.

x	0	1	2	3	4	5	6	7 or more
Frequency	12.2	27.0	28.5	19.0	9	3.2	0.9	0.2

Table 2

Stating your hypotheses clearly, use a 5% level of significance to test the manager's claim. **(7)**

<u>Solutions</u>

1.

<table>
<tr><td>
<table>
<tr><td>$\boldsymbol{x}$</td><td>0</td><td>1</td><td>2</td><td>3</td><td>≥4</td></tr>
<tr><td>$\boldsymbol{O_i}$</td><td>17</td><td>31</td><td>19</td><td>14</td><td>19</td></tr>
<tr><td>$\boldsymbol{E_i}$</td><td>12.2</td><td>27.0</td><td>28.5</td><td>19.0</td><td>13.3</td></tr>
<tr><td>$\boldsymbol{\frac{(O-E)^2}{E}}$</td><td>1.89</td><td>0.59</td><td>3.17</td><td>1.32</td><td>2.44</td></tr>
</table>
</td><td>M1</td></tr>
<tr><td>$\sum \frac{(O-E)^2}{E} = 9.41$ AWRT 9.4</td><td>M1A1</td></tr>
<tr><td>$v = 5 - 2 = 3$</td><td>B1</td></tr>
<tr><td>χ^2_3 (5%) = 7.815</td><td>B1</td></tr>
<tr><td>H_0: Binomial distribution is a good/suitable model/fit
[Condone: B(20, 0.1) is...]

H1: Binomial distribution is not a suitable model

(for hypotheses) allow just “$X \sim$ B(20, 0.1)” for null etc.</td><td>both B1</td></tr>
<tr><td>(Significant result) Binomial distribution is not a suitable model</td><td>A1</td></tr>
</table>

Topic: Chi-Squared Tests (2)

Chapter Reference: Further Statistics 1, Chapter 6

10 minutes

1. An area of grass was sampled by placing a 1 m × 1 m square randomly in 100 places. The numbers of daisies in each of the squares were counted. It was decided that the resulting data could be modelled by a Poisson distribution with mean 2. The expected frequencies were calculated using the model.

The following table shows the observed and expected frequencies.

Number of daisies	**Observed frequency**	**Expected frequency**
0	8	13.53
1	32	27.07
2	27	r
3	18	s
4	10	9.02
5	3	3.61
6	1	1.20
7	0	0.34
≥8	1	0.12

a. Find values for r and s. **(2)**

b. Using a 5% significance level, test whether or not this Poisson model is suitable. State your hypotheses clearly. **(7)**

Solutions

1a.

r = 27.07	**A1**
s = 18.04	**A1**

1b.

H_0: A Poisson model Po(2) is a suitable model. H_1: A Poisson model Po(2) is not a suitable model.	**Both B1**
Amalgamate data	**M1**
$\sum \frac{(O-E)^2}{E}$ = 3.28 (awrt)	**M1A1**
v = = 6 – 1 = 5	**B1**
χ^2_5 (5 %) = 11.070 (follow through their degrees of freedom)	**B1**
3.25 < 11.070 There is insufficient evidence to reject H_0, Po(2) is a suitable model.	**A1**

Topic: Chi-Squared Tests and Contingency Tables (3)

Chapter Reference: Further Statistics 1, Chapter 6

10 minutes

1. Year 12 students at Coron Academy choose to participate in one of four sports during the Spring term, The students' choices are summarised in the table.

	Squash	**Badminton**	**Archery**	**Hockey**	**Total**
Male	5	16	30	19	70
Female	4	20	33	53	110
Total	9	36	63	72	180

Use a χ^2 test, at the 5% level of significance, to determine whether the choice of sport is independent of gender. **(10)**

Solutions

1.

H_0: Choice independent of gender	**B1**

	Squash	**Badminton**	**Archery**	**Hockey**
Male	5/3.5	16/14	30/24.5	19/28
Female	4/3.5	20/22	33/38.5	53/44

M1

Combine Squash and Badminton	**M1**

	Squash & Badminton	**Archery**	**Hockey**
Male	21/17.5	30/24.5	19/28
Female	24/27.5	33/38.5	53/44

M1
M1

χ^2 values

	Squash & Badminton	**Archery**	**Hockey**
Male	0.7000	1.2347	2.8928
Female	0.4455	0.7857	1.8409

M1

$\chi^2_{\text{calc}} = 7.90$ (7.8 to 7.9)	**A1**
$v = 2$	**B1**
$\chi^2_{5\%}(2) = 5.991$	**B1**
Reject H_0 Sufficient evidence, at the 1% level of significance, to support an association between the choice of sport and gender	**A1**

Further Maths A-Level Starter Activity

Topic: Chi-Squared Tests and Geometric Distribution (4)

Chapter Reference: Further Statistics 1, Chapter 6

10 minutes

1. The following table shows observed values for what is thought to be a geometric distribution with $p = 0.5$.

x	1	2	3	4	5	6	**Total**
Observed frequency O_x	56	27	13	3	0	1	100

Calculate the expected frequencies and, using a 5% significance level, conduct a good-of-fit test. **(6)**

Solutions

1.

<table>
<tr><td>H_0: X ~ Geo(0.5) is a suitable model
H_1: X ~ Geo(0.5) is not a suitable model</td><td>B1</td></tr>
<tr><td><table>
<tr><td>x</td><td>1</td><td>2</td><td>≥3</td><td>Total</td></tr>
<tr><td>Expected frequency E_x</td><td>50</td><td>25</td><td>25</td><td>100</td></tr>
</table></td><td>M1</td></tr>
<tr><td><table>
<tr><td>x</td><td>1</td><td>2</td><td>≥3</td><td>Total</td></tr>
<tr><td>$\frac{(O_x - E_x)^2}{E_x}$</td><td>0.72</td><td>0.16</td><td>2.56</td><td>3.44</td></tr>
</table>
$X^2 = 3.44$</td><td>M1</td></tr>
<tr><td>$v = 3 - 1 = 2$</td><td>B1</td></tr>
<tr><td>$\chi_2^2(5\%) = 5.991$</td><td>B1</td></tr>
<tr><td>Since 5.991 > 3.44 there is insufficient evidence to reject H_0 at the 5% level
X ~ Geo(0.5) is a suitable model for the data.</td><td>A1</td></tr>
</table>

Topic: Chi-Squared Tests and Geometric Distribution (5)

Chapter Reference: Further Statistics 1, Chapter 6

10 minutes

1. Eli is practising the water bottle flip for a YouTube video. He records how many attempts he needs to correctly flip the water bottle on 100 separate occasions.

Attempts	1	2	3	4	5	**Total**
Frequency	34	22	14	18	12	100

a. Using the observed frequencies, find an estimate for p to 3.d.p. **(2)**

b. Conduct a goodness-of-fit test at the 2.5% significance level, and say whether a geometric random variable is a good model for the data. **(6)**

Solutions

1a.

$p = \frac{1}{\bar{x}}$ $\bar{x} = \frac{\sum fx}{\sum f} = \frac{(34\times1) + (22\times2) + (14\times3) + (18\times4) + (12\times5)}{100} = 2.52$	**M1**
$p = \frac{1}{2.52} \approx 0.397$	**A1**

1b.

H_0: A geometric distribution is a suitable model H_1: A geometric distribution is not a suitable model	**B1**

x	1	2	3	4	≥5	**Total**
Expected frequency E_i	39.7	23.9	14.4	8.7	13.3	100

M1

x	1	2	3	4	≥5	**Total**
$\frac{(O_i - E_i)^2}{E_i}$	0.818	0.151	0.011	9.941	0.127	11.048

$X^2 = 11.048$

M1

$v = 5 - 1 = 4$	**B1**
$\chi_4^2(2.5\%) = 11.143$	**B1**
Since 11.143 > 11.048 there is insufficient evidence to reject H_0 at the 2.5% level. Geometric distribution is a suitable model for the data.	**A1**