Chi squared test 2 [62 marks]

1. [Maximum mark: 18] EXM.1.AHL.TZ0.56

A zoologist believes that the number of eggs laid in the Spring by female birds of a certain breed follows a Poisson law. She observes 100 birds during this period and she produces the following table.

Number of eggs laid	Frequency
0	10
1	19
2	34
3	23
4	10
5	4

(a) Calculate the mean number of eggs laid by these birds. [2]

Markscheme

Mean $= \frac{1\times19+2\times34+\ldots+5\times4}{100}$ ***(M1)***

$= 2.16$ ***A1 N2***

[2 marks]

The zoologist wishes to determine whether or not a Poisson law provides a suitable model.

(b.i) Write down appropriate hypotheses. [2]

Markscheme

H_0 : Poisson law provides a suitable model ***A1***

H_1 : Poisson law does not provide a suitable model ***A1***

[2 marks]

(b.ii) Carry out a test at the 1% significance level, and state your conclusion. [14]

Markscheme

The expected frequencies are

Number of eggs	Observed frequency	Expected frequency
0	10	11.533
1	19	24.910
2	34	26.903
3	23	19.370
4	10	10.460
5 or more	4	6.824

A1A1A1A1A1A1

Note: Accept expected frequencies rounded to a minimum of three significant figures.

$\chi^2 = \frac{(10-11.533)^2}{11.533} + \ldots + \frac{(4-6.824)^2}{6.824}$ ***(M1)(A2)***

$= 5.35$ (accept 5.33 and 5.34) ***A2***

$v = 4$ (6 cells – 2 restrictions) ***A1***

Note: If candidates have combined rows allow FT on their value of v.

Critical value $\chi^2 = 13.277$

Because 5.35 < 13.277, the Poisson law does provide a suitable model. ***R1 N0***

[14 marks]

2. [Maximum mark: 10] EXM.1.AHL.TZ0.55

Eggs at a farm are sold in boxes of six. Each egg is either brown or white. The owner believes that the number of brown eggs in a box can be modelled by a binomial distribution. He examines 100 boxes and obtains the following data.

Number of brown eggs in a box	Frequency
0	10
1	29
2	31
3	18
4	8
5	3
6	1

(a.i) Calculate the mean number of brown eggs in a box. [1]

Markscheme

Note: *Candidates may obtain slightly different numerical answers depending on the calculator and approach used. Use discretion in marking.*

Mean $= \frac{1\times 29+\ldots+6\times 1}{100} = 1.98$ ***(A1)***

[1 mark]

(a.ii) Hence estimate p, the probability that a randomly chosen egg is brown. [1]

Markscheme

Note: *Candidates may obtain slightly different numerical answers depending on the calculator and approach used. Use discretion in marking.*

$\widehat{p} = \frac{1.98}{6} = 0.33$ ***(A1)***

[1 mark]

(b) By calculating an appropriate χ^2 statistic, test, at the 5% significance level, whether or not the binomial distribution gives a good fit to these data. [8]

Markscheme

Note: *Candidates may obtain slightly different numerical answers depending on the calculator and approach used. Use discretion in marking.*

The calculated values are

f_0	f_e	$(f_0 - f_e)^2$	
10	9.046	0.910	
29	26.732	5.14	***(M1)***
31	32.917	3.675	***(A1)***
18	21.617	13.083	***(A1)***
12	9.688	5.345	***(A1)***

Note: Award ***(M1)*** for the attempt to calculate expected values, ***(A1)*** for correct expected values, ***(A1)*** for correct $(f_0 - f_e)^2$ values, ***(A1)*** for combining cells.

$\chi^2 = \frac{0.910}{9.046} + \ldots + \frac{5.345}{9.688} = 1.56$ ***(A1)***

OR

$\chi^2 = 1.56$ ***(G5)***

Degrees of freedom = 3; Critical value = 7.815

(or p-value = 0.668 (or 0.669)) ***(A1)(A1)***

We conclude that the binomial distribution does provide a good fit. ***(R1)***

[8 marks]

3. [Maximum mark: 12] EXM.2.AHL.TZ0.29

(a) A horse breeder records the number of births for each of 100 horses during the past eight years. The results are summarized in the following table:

Number of births	0	1	2	3	4	5	6
Frequency	1	5	26	37	18	12	1

Stating null and alternative hypotheses carry out an appropriate test at the 5% significance level to decide whether the results can be modelled by B (6, 0.5). [10]

Markscheme

METHOD 1

H_0: distribution is B(6, 0.5); H_1: distribution is not B(6, 0.5) ***A1***

	0	1	2	3	4	5	6
Observed frequency	1	5	26	37	18	12	1
Expected frequency	$\frac{25}{16}$ = 1.5625	$\frac{150}{16}$ = 9.375	$\frac{375}{16}$ = 23.4375	$\frac{500}{16}$ = 31.25	$\frac{375}{16}$ = 23.4375	$\frac{150}{16}$ = 9.375	$\frac{25}{16}$ = 1.5625

$\left(E_0 = 100(0.5)^6 = \frac{25}{16} = 0.015625\right)$ ***A3***

Combining the first two columns and the last two columns: ***A1***

$\chi^2 = \sum \frac{O^2}{E} - \sum E$

$= \frac{6^2}{\left(\frac{175}{16}\right)} + \frac{26^2}{\left(\frac{375}{16}\right)} + \frac{37^2}{\left(\frac{500}{16}\right)} + \frac{18^2}{\left(\frac{375}{16}\right)} + \frac{13^2}{\left(\frac{175}{16}\right)} - 100$ ***(M1)***

= 5.22 ***A1***

v = 4, so critical value of $\chi^2_{5\%} = 9.488$ ***A1A1***

Since 5.22 < 9.488 the result is not significant and we accept H_0 ***R1***

METHOD 2

H_0: distribution is B(6, 0.5); H_1: distribution is not B(6, 0.5) ***A1***

By GDC, $p = 0.266$ ***A8***

Since 0.266 > 0.05 the result is not significant and we accept H_0 ***R1***

[10 marks]

(b) Without doing any further calculations, explain briefly how you would carry out a test, at the 5% significance level, to decide if the data can be modelled by B(6, p), where p is unspecified. [2]

Markscheme

Estimate p from the data which would entail the loss of one degree of freedom ***A1A1***

[2 marks]

4. [Maximum mark: 12] EXM.2.AHL.TZ0.24

The hens on a farm lay either white or brown eggs. The eggs are put into boxes of six. The farmer claims that the number of brown eggs in a box can be modelled by the binomial distribution, B(6, p). By inspecting the contents of 150 boxes of eggs she obtains the following data.

Number of brown eggs	0	1	2	3	4	5	6
Number of boxes	7	32	35	50	22	4	0

(a) Show that this data leads to an estimated value of $p = 0.4$. [1]

Markscheme

from the sample, the probability of a brown egg is

$\frac{0\times7+1\times32+\ldots}{6\times150} = \frac{360}{900} = 0.4$ ***A1***

$p = 0.4$ ***AG***

[1 mark]

(b) Stating null and alternative hypotheses, carry out an appropriate test at the 5 % level to decide whether the farmer's claim can be justified. [11]

Markscheme

if the data can be modelled by a binomial distribution with $p = 0.4$, the expected frequencies of boxes are given in the table

Number of brown eggs	0	1	2	3	4	5, 6
Number of boxes	7	32	35	50	22	4
Number of boxes	7.0	28.0	46.7	41.5	20.7	6.1

A3

Notes: Deduct one mark for each error or omission.
Accept any rounding to at least one decimal place.

null hypothesis: the distribution is binomial ***A1***

alternative hypothesis: the distribution is not binomial ***A1***

for a chi-squared test the last two columns should be combined ***R1***

Number of brown eggs	0	1	2	3	4	5	6
Number of boxes	7	32	35	50	22	4	0
Number of boxes	7.0	28.0	46.7	41.5	20.7	5.5	0.6

$\chi^2_{\text{calc}} = \frac{(7-7)^2}{7} + \frac{(32-28)^2}{28} + \ldots = 6.05$ (Accept 6.06) ***(M1)A1***

degrees of freedom = 4 ***A1***

critical value = 9.488 ***A1***

Or use of p-value

we conclude that the farmer's claim can be justified ***R1***

[11 marks]

5. [Maximum mark: 10] EXM.2.AHL.TZ0.26

Scientists have developed a type of corn whose protein quality may help chickens gain weight faster than the present type used. To test this new type, 20 one-day-old chicks were fed a ration that contained the new corn while another control group of 20 chicks was fed the ordinary corn. The data below gives the weight gains in grams, for each group after three weeks.

Ordinary corn (Group A)				New corn (Group B)			
380	321	366	356	361	447	401	375
283	349	402	462	434	403	393	426
356	410	329	399	406	318	467	407
350	384	316	272	427	420	470	392
345	455	360	431	430	339	410	326

(a) The data from the two samples above are combined to form a single set of data. The following frequency table gives the observed frequencies for the combined sample. The data has been divided into five intervals.

Weight gain	Observed
271–310	2
311–350	9
351–390	8
391–430	15
431–470	6

Test, at the 5% level, whether the combined data can be considered to be a sample from a normal population with a mean of 380. [10]

Markscheme

This is a χ^2 goodness-of-fit test.

To finish the table, the frequencies of the respective cells have to be calculated. Since the standard deviation is not given, it has to be estimated using the data itself. $s = 49.59$, *eg* the third expected frequency is $40 \times 0.308 = 12.32$, since $P(350.5 < W < 390.5) = 0.3078...$

The table of observed and expected frequencies is:

Amount of weight gain	Observed	Expected
271 – 310	2	3.22
311 – 350	9	7.82
351 – 390	8	12.32
391 – 430	15	10.48
431 – 470	6	6.17

(M1)(A2)

Since the first expected frequency is 3.22, we combine the two cells, so that the first two rows become one row, that is,

271 – 350	11	11.04

(M1)

Number of degrees of freedom is 4 – 1 – 1 = 2 ***(C1)***

H_0: The distribution is normal with mean 380

H_1: The distribution is not normal with mean 380 ***(A1)***

The test statistic is

$$\chi^2_{calc} = \sum \frac{(f_e - f_0)^2}{f_e} = \frac{(11-11.04)^2}{11.04} + \frac{(8-12.32)^2}{12.32} + \frac{(15-10.48)^2}{10.48} + \frac{(6-6.17)^2}{6.17}$$

= 3.469 ***(A1)***

With 2 degrees of freedom, the critical number is χ^2 = 5.99 ***(A2)***

So, we do not have enough evidence to reject the null hypothesis. Therefore, there is no evidence to say that the distribution is not normal with mean 380. ***(R1)***

[10 marks]