Chi Squared Tests for Goodness of Fit (From Edexcel 6691)

Q1, (Jun 2007, Q4)

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

(a) Estimate the proportion of defective items from the production line.

(2)

The manager claimed that the number of defective items in a sample of 20 can be modelled by a binomial distribution. He used the answer in part (a) to calculate the expected frequencies given in Table 2.

x	0	1	2	3	4	5	6	7 or more
Expected frequency	12.2	27.0	r	19.0	s	3.2	0.9	0.2

Table 2

(b) Find the value of r and the value of s giving your answers to 1 decimal place.

(3)

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

(7)

(d) Explain what the analysis in part (c) tells the manager about the occurrence of defective items from this production line.

(1)

Q2, (Jun 2010, Q6)

A total of 228 items are collected from an archaeological site. The distance from the centre of the site is recorded for each item. The results are summarised in the table below.

Distance from the centre of the site (m)	0–1	1–2	2–4	4–6	6–9	9–12
Number of items	22	15	44	37	52	58

Test, at the 5% level of significance, whether or not the data can be modelled by a continuous uniform distribution. State your hypotheses clearly.

(12)

Q3, (Jun 2008, Q6)

Ten cuttings were taken from each of 100 randomly selected garden plants. The numbers of cuttings that did not grow were recorded.

The results are as follows

No. of cuttings which did not grow	0	1	2	3	4	5	6	7	8, 9 or 10
Frequency	11	21	30	20	12	3	2	1	0

(a) Show that the probability of a randomly selected cutting, from this sample, not growing is 0.223

(2)

A gardener believes that a binomial distribution might provide a good model for the number of cuttings, out of 10, that do not grow.

He uses a binomial distribution, with the probability 0.2 of a cutting not growing. The calculated expected frequencies are as follows

No. of cuttings which did not grow	0	1	2	3	4	5 or more
Expected frequency	r	26.84	s	20.13	8.81	t

(b) Find the values of r, s and t.

(4)

(c) State clearly the hypotheses required to test whether or not this binomial distribution is a suitable model for these data.

(2)

The test statistic for the test is 4.17 and the number of degrees of freedom used is 4.

(d) Explain fully why there are 4 degrees of freedom.

(2)

(e) Stating clearly the critical value used, carry out the test using a 5% level of significance.

(3)

Q4. (Jun 2009, Q5)

The number of goals scored by a football team is recorded for 100 games. The results are summarised in Table 1 below.

Number of goals	Frequency
0	40
1	33
2	14
3	8
4	5

Table 1

(a) Calculate the mean number of goals scored per game.

(2)

The manager claimed that the number of goals scored per match follows a Poisson distribution. He used the answer in part (a) to calculate the expected frequencies given in Table 2.

Number of goals	Expected Frequency
0	34.994
1	r
2	s
3	6.752
$\geqslant 4$	2.221

Table 2

(b) Find the value of r and the value of s giving your answers to 3 decimal places.

(3)

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

(7)

Q5, (Jun 2011, Q5)

The number of hurricanes per year in a particular region was recorded over 80 years. The results are summarised in Table 1 below.

No of hurricanes, h	0	1	2	3	4	5	6	7
Frequency	0	2	5	17	20	12	12	12

Table 1

(a) Write down two assumptions that will support modelling the number of hurricanes per year by a Poisson distribution.

(2)

(b) Show that the mean number of hurricanes per year from Table 1 is 4.4875

(2)

(c) Use the answer in part (b) to calculate the expected frequencies r and s given in Table 2 below to 2 decimal places.

(3)

h	0	1	2	3	4	5	6	7 or more
Expected frequency	0.90	4.04	r	13.55	s	13.65	10.21	13.39

Table 2

(d) Test, at the 5% level of significance, whether or not the data can be modelled by a Poisson distribution. State your hypotheses clearly.

(6)

Q6, (Jun 2012, Q6)

A total of 100 random samples of 6 items are selected from a production line in a factory and the number of defective items in each sample is recorded. The results are summarised in the table below.

Number of defective items	0	1	2	3	4	5	6
Number of samples	6	16	20	23	17	10	8

(a) Show that the mean number of defective items per sample is 2.91

(2)

A factory manager suggests that the data can be modelled by a binomial distribution with $n = 6$. He uses the mean from the sample above and calculates expected frequencies as shown in the table below.

Number of defective items	0	1	2	3	4	5	6
Expected frequency	1.87	10.54	24.82	a	22.01	8.29	b

(b) Calculate the value of a and the value of b giving your answers to 2 decimal places.

(4)

(c) Test, at the 5% level, whether or not the binomial distribution is a suitable model for the number of defective items in samples of 6 items.
State your hypotheses clearly.

(8)

Q7, (Jun 2014, Q5)

A research station is doing some work on the germination of a new variety of genetically modified wheat.

They planted 120 rows containing 7 seeds in each row.

The number of seeds germinating in each row was recorded. The results are as follows

Number of seeds germinating in each row	0	1	2	3	4	5	6	7
Observed number of rows	2	6	11	19	25	32	16	9

(a) Write down two reasons why a binomial distribution may be a suitable model.

(2)

(b) Show that the probability of a randomly selected seed from this sample germinating is 0.6

(2)

The research station used a binomial distribution with probability 0.6 of a seed germinating. The expected frequencies were calculated to 2 decimal places. The results are as follows

Number of seeds germinating in each row	0	1	2	3	4	5	6	7
Expected number of rows	0.20	2.06	s	23.22	t	31.35	15.68	3.36

(c) Find the value of s and the value of t.

(2)

(d) Stating your hypotheses clearly, test, at the 1% level of significance, whether or not the data can be modelled by a binomial distribution.

(7)

Q8, (Jun 2016, Q6)

An airport manager carries out a survey of families and their luggage. Each family is allowed to check in a maximum of 4 suitcases. She observes 50 families at the check-in desk and counts the total number of suitcases each family checks in. The data are summarised in the table below.

Number of suitcases	0	1	2	3	4
Frequency	6	25	12	6	1

The manager claims that the data can be modelled by a binomial distribution with $p = 0.3$

(a) Test the manager's claim at the 5% level of significance. State your hypotheses clearly.
Show your working clearly and give your expected frequencies to 2 decimal places.

(8)

The manager also carries out a survey of the time taken by passengers to check in. She records the number of passengers that check in during each of 100 five-minute intervals.

The manager makes a new claim that these data can be modelled by a Poisson distribution. She calculates the expected frequencies given in the table below.

Number of passengers	0	1	2	3	4	5 or more
Observed frequency	5	40	31	18	6	0
Expected frequency	16.53	29.75	r	s	7.23	3.64

(b) Find the value of r and the value of s giving your answers to 2 decimal places.

(3)

(c) Stating your hypotheses clearly, use a 1% level of significance to test the manager's new claim.

(6)