FS1 (Year 1) – Chi-Squared Goodness Of Fit Tests Exam Questions (Edexcel)

Q1.

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)

(Total for question = 17 marks)

(Q04 6691/01, June 2016)

Q2.

A researcher is investigating the distribution of orchids in a field. He believes that the Poisson distribution with a mean of 1.75 may be a good model for the number of orchids in each square metre. He randomly selects 150 non-overlapping areas, each of one square metre, and counts the number of orchids present in each square.

The results are recorded in the table below.

Number of orchids in each square metre	0	1	2	3	4	5	6
Number of squares	30	42	35	26	11	6	0

He calculates the **expected** frequencies as follows

Number of orchids in each square metre	0	1	2	3	4	5	More than 5
Number of squares	26.07	45.62	39.91	23.28	10.19	3.57	r

(a) Find the value of r giving your answer to 2 decimal places.

(1)

The researcher will test, at the 5% level of significance, whether or not the data can be modelled by a Poisson distribution with mean 1.75

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

(1)

The test statistic for this test is 2.0 and the number of degrees of freedom to be used is 4

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

(2)

(d) Stating your critical value clearly, determine whether or not these data support the researcher's belief.

(2)

The researcher works in another field where the number of orchids in each square metre is known to have a Poisson distribution with mean 1.5

He randomly selects 200 non-overlapping areas, each of one square metre, in this second field, and counts the number of orchids present in each square.

(e) Using a Poisson approximation, show that the probability that he finds at least one square with exactly 6 orchids in it is 0.506 to 3 decimal places.

(4)

(Total for question = 10 marks)

(Q01 8FM0/23, June 2018)

Q3.

Bags of £1 coins are paid into a bank. Each bag contains 20 coins.

The bank manager believes that 5% of the £1 coins paid into the bank are fakes. He decides to use the distribution $X \sim B(20, 0.05)$ to model the random variable X, the number of fake £1 coins in each bag.

(a) State the assumptions necessary for the binomial distribution to be an appropriate model in this case.

(2)

The bank manager checks a random sample of 150 bags of £1 coins and records the number of fake coins found in each bag. His results are summarised in Table 1.

Number of fake coins in each bag	0	1	2	3	4 or more
Observed frequency	43	62	26	13	6
Expected frequency	53.8	56.6	r	8.9	s

Table 1

(b) Calculate the values of r and s, giving your answers to 1 decimal place.

(3)

(c) Carry out a hypothesis test, at the 5% significance level, to see if the data supports the bank manager's statistical model. State your hypotheses clearly.

(7)

The assistant manager thinks that a binomial distribution is a good model but suggests that the proportion of fake coins is higher than 5%. She calculates the actual proportion of fake coins in the sample and uses this value to carry out a new hypothesis test on the data. Her expected frequencies are shown in Table 2.

Number of fake coins in each bag	0	1	2	3	4 or more
Observed frequency	43	62	26	13	6
Expected frequency	44.5	55.7	33.2	12.5	4.1

Table 2

(d) Explain why there are 2 degrees of freedom in this case.

(2)

(e) Given that she obtains a χ^2 test statistic of 2.67, test the assistant manager's hypothesis that the binomial distribution is a good model for the number of fake coins in each bag. Use a 5% level of significance and state your hypotheses clearly.

(3)

(Total 17 marks)

(Q03 6691/01/R, June 2014)

Q4.

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)

(Total 13 marks)

(Q03 6691/01, June 2014)

Q5.

The discrete random variable X follows a Poisson distribution with mean 1.4

(a) Write down the value of

(i) $P(X = 1)$
(ii) $P(X \leq 4)$

(2)

The manager of a bank recorded the number of mortgages approved each week over a 40 week period.

Number of mortgages approved	0	1	2	3	4	5	6
Frequency	10	16	7	4	2	0	1

(b) Show that the mean number of mortgages approved over the 40 week period is 1.4

(1)

The bank manager believes that the Poisson distribution may be a good model for the number of mortgages approved each week.

She uses a Poisson distribution with a mean of 1.4 to calculate expected frequencies as follows.

Number of mortgages approved	0	1	2	3	4	5 or more
Expected frequency	9.86	r	9.67	4.51	1.58	s

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

(2)

The bank manager will test, at the 5% level of significance, whether or not the data can be modelled by a Poisson distribution.

(d) Calculate the test statistic and state the conclusion for this test. State clearly the degrees of freedom and the hypotheses used in the test.

(6)

(Total for question = 11 marks)

(Q04 8FM0/2G/sA, Specimen papers)

Q6.

A spinner used for a game is designed to give scores with the following probabilities

Score	1	2	3	4	6
Probability	$\frac{3}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{2}{5}$	$\frac{1}{10}$

The spinner is spun 80 times and the results are as follows

Score	1	2	3	4	6
Frequency	15	4	12	41	8

Test, at the 10% level of significance, whether or not the spinner is giving scores as it is designed to do. Show your working and state your hypotheses clearly.

(Total for question = 7 marks)

(Q02 8FM0/23, June 2019)

Q7.

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)

[Question Continued On Next Page]

(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)

(Total 13 marks)

(Q04 6691/01, June 2008)

Q8.

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)

(Total 13 marks)

(Q05 6691/01, June 2007)

Q9.

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)

(Total 14 marks)

(Q03 6691/01, June 2012)

Q10.

In a game, a coin is spun 5 times and the number of heads obtained is recorded.
Tao suggests playing the game 20 times and carrying out a chi-squared test to investigate whether the coin might be biased.

(a) Explain why playing the game only 20 times may cause problems when carrying out the test.

(1)

Chris decides to play the game 500 times. The results are as follows

Number of heads	0	1	2	3	4	5
Observed frequency	2	27	93	181	146	51

Chris decides to test whether or not the data can be modelled by a binomial distribution, with the probability of a head on each spin being 0.6

She calculates the expected frequencies, to 2 decimal places, as follows

Number of heads	0	1	2	3	4	5
Expected frequency	5.12	38.40	115.20	172.80	129.60	38.88

(b) State the number of degrees of freedom in Chris' test, giving a reason for your answer.

(1)

(c) Carry out the test at the 5% level of significance.
You should state your hypotheses, test statistic, critical value and conclusion clearly.

(5)

(d) Showing your working, find an alternative model which would better fit Chris' data.

(2)

(Total for question = 9 marks)

(Q03 8FM0/23, June 2022)