2.6 Further Statistics: Hypothesis Testing and Chi-Squared Depth

Total Marks

EDEXCEL 9FM0 · Calculator allowed · about 230 minutes · Write-in edition

Name: _____ Date: ____ / ____ / _____

Answer ALL questions. Show all your working.

Key facts: chi-squared tests, in more depth

Original content written for Revision Library.

A chi-squared test compares observed frequencies O with the expected frequencies E predicted by a model, using X^2 = sum of (O-E)^2/E over all classes or cells, after any pooling. For a goodness-of-fit test to a fully specified distribution, df = (number of classes) - 1; if a parameter is estimated from the same sample (for example a Poisson mean, a binomial p, or a geometric p), one further degree of freedom is lost for each parameter estimated, because the estimation process artificially improves the apparent fit. For continuous data grouped into classes, expected frequencies are found from probabilities calculated using the standard Normal distribution, via z = (x - mu)/sigma and the function Phi. For an r by c contingency table testing independence between two categorical variables, E = (row total * column total)/grand total (this follows directly from the multiplication rule P(row and column) = P(row) * P(column) applied to the sample), and df = (r-1)(c-1). Classes or cells with an expected frequency below 5 should be pooled with a neighbouring class or column before the test statistic is calculated, and the degrees of freedom adjusted accordingly. Yates' continuity correction, X^2 = sum of (|O-E| - 0.5)^2/E, is applied only when df = 1 (every 2 by 2 contingency table), since this is the only case where the discrete test statistic is poorly approximated by the continuous chi-squared distribution. The signed standardised residual for a cell, e = (O-E)/sqrt(E), shows the direction as well as the size of a departure from the model, and the sum of its squares reproduces X^2. Because X^2 is never negative, only large (upper-tail) values ever provide evidence against H0; the calculated statistic is compared with the critical value from chi-squared tables at the chosen significance level and the stated degrees of freedom, and H0 is rejected if the statistic exceeds this value. As with any hypothesis test, the significance level equals P(Type I error), reducing it increases P(Type II error) for a fixed sample size, and increasing the sample size generally increases the power of the test; a chi-squared statistic also scales with sample size, so a very large sample can make even a small, unimportant deviation from a model 'statistically significant'. Combining ('pooling') data from distinct groups before testing can also mask or even reverse a genuine association found within each group, an effect known as Simpson's paradox.

1 Recall that two discrete random variables X and Y are said to be independent if P(X = x and Y = y) = P(X = x) * P(Y = y) for every possible pair of values x and y (the multiplication rule used, for example, when finding E(XY) for independent variables in FP.FS1).

(a) Explain how the null hypothesis of a chi-squared test for association between two categorical variables is connected to this definition of independence. **(2)**

(b) For a general 2 by 2 contingency table with row totals R1 and R2, column totals C1 and C2, and grand total n, the sample estimates of the marginal probabilities are P(row 1) = R1/n and P(column 1) = C1/n. Assuming independence, so that P(row 1 and column 1) = P(row 1) * P(column 1), show that the expected frequency for the (row 1, column 1) cell is exactly R1*C1/n, the usual formula E = (row total * column total) / grand total. **(2)**

(c) A student says: 'If the chi-squared test statistic works out to be exactly 0, the two variables must be perfectly independent.' Comment on whether this statement is accurate. **(2)**

(Total for Question 1 is 6 marks)

2 When one or more parameters of a hypothesised distribution are estimated from the very same sample that is then used for the goodness-of-fit test, an extra degree of freedom is lost for each parameter estimated.

(a) Explain, in your own words, why estimating a parameter from the sample data (rather than using a value fixed in advance) tends to make the fitted distribution match that sample more closely than it otherwise would, and hence why the degrees of freedom must be reduced to compensate. **(3)**

(b) A goodness-of-fit test is to be carried out for a Normal distribution $N(\mu, \sigma^2)$ using 6 weight classes. State the number of degrees of freedom in each of the following cases: (i) both μ and σ are specified in advance; (ii) μ is specified in advance but σ is estimated from the sample; (iii) both μ and σ are estimated from the sample. **(3)**

(Total for Question 2 is 6 marks)

3 A tutoring company in Leicester runs an optional Saturday revision club before a Year 11 mock exam. Of 150 students entered for the mock, the company records whether each student attended the Saturday club and whether they passed the mock. The results are given below.

Attended club: passed 60, did not pass 20, row total 80.
Did not attend: passed 30, did not pass 40, row total 70.
Column totals: passed 90, did not pass 60, grand total 150.

(a) State suitable null and alternative hypotheses for this test. **(2)**

(b) Calculate the expected frequency for each of the four cells. **(2)**

(c) Calculate the value of the chi-squared test statistic. **(3)**

(d) State the number of degrees of freedom for this test. Using the critical values chi-squared$_1$(0.05) = 3.841 and chi-squared$_1$(0.005) = 7.879, test at the 5% significance level and comment on the strength of the evidence found. **(2)**

(e) Yates' continuity correction is sometimes applied to 2 by 2 contingency tables. State a condition under which it would usually be recommended, and explain why, even without applying it here, you can be confident it would not change the conclusion of this test. **(2)**

(f) Explain why Yates' continuity correction is only ever applied to chi-squared tests with 1 degree of freedom (that is, 2 by 2 contingency tables), and not to larger contingency tables. **(2)**

(Total for Question 3 is 13 marks)

4 A market researcher carries out a chi-squared goodness-of-fit test on a random sample of 50 survey responses split across 4 equally likely categories under H0 (so each expected frequency is 12.5), obtaining the observed frequencies 8, 18, 14 and 10.

(a) Calculate the value of the chi-squared test statistic for this sample of 50 responses, and test at the 5% significance level using the critical value chi-squared$_3$(0.05) = 7.815. **(3)**

(b) A second researcher repeats the same survey with the same proportional pattern of responses, but with a sample 10 times as large, so every observed and expected frequency above is exactly 10 times as large (that is, O' = 10*O and E' = 10*E in each category). Without recalculating from scratch, show algebraically that the new chi-squared test statistic must be exactly 10 times the value found in part (a), and hence state its value. **(3)**

(c) Test the result from part (b) at the 5% significance level, and comment on what this reveals about the relationship between sample size and the outcome of a chi-squared test, even when the underlying pattern of proportions is unchanged. **(3)**

(Total for Question 4 is 9 marks)

5 A local council carries out a chi-squared test for association at the 5% significance level between the area of a town (North, South, East) and residents' preferred recycling collection frequency (Weekly, Fortnightly), based on a survey of 300 residents. The test gives a test statistic of 8.24 with 2 degrees of freedom; the critical value at the 5% level is 5.991.

(a) State the conclusion of the test, in context. **(2)**

(b) Explain what is meant by a Type I error in the context of this test. **(2)**

(c) State the probability of a Type I error occurring in this test. **(1)**

(d) The council wants to reduce the probability of a Type I error occurring. State one way it could achieve this, and explain the effect this would have on the probability of a Type II error. **(2)**

(e) Explain why increasing the sample size, while keeping the significance level fixed, generally reduces the probability of a Type II error. **(2)**

(f) State one condition, referring to this context, that must hold for the chi-squared test used here to be valid. **(1)**

(Total for Question 5 is 10 marks)

---

6 A researcher counts the number of events per interval over 119 intervals: 0 events 24 times, 1 event 37 times, 2 events 32 times, and 3 or more events 26 times. It is suggested that the number of events per interval follows a Poisson distribution. Taking the mean of the data to be 1.504 (treating '3 or more' as 3), test the suggestion at the 5% level.

(a) Calculate the expected frequencies for a Poisson distribution with the given mean, using the categories 0, 1, 2 and 3 or more. **(3)**

(b) Calculate the test statistic and the degrees of freedom, and state the conclusion at the 5% level of significance, given that the critical value is 5.991. **(4)**

(Total for Question 6 is 7 marks)

7 A fair coin-like device is claimed to land on 'success' with probability 0.25. A tester believes the probability of success is actually lower. Let X be the number of trials needed to get the first success, so under the claim X ~ Geo(0.25). The first success occurs on trial 13.

(a) State suitable hypotheses for the test, in terms of p. **(2)**

(b) Find $P(X \geq 13)$ assuming H0 is true, and test the tester's belief at the 5% level. **(3)**

(c) Find the critical region for the test at the 5% level, and the probability of a Type II error if the true value of p is 0.1, to 3 significant figures. **(4)**

(Total for Question 7 is 9 marks)

8 A researcher counts the number of events per interval over 149 intervals: 0 events 56 times, 1 event 54 times, 2 events 30 times, and 3 or more events 9 times. It is suggested that the number of events per interval follows a Poisson distribution. Taking the mean of the data to be 0.946 (treating '3 or more' as 3), test the suggestion at the 5% level.

(a) Calculate the expected frequencies for a Poisson distribution with the given mean, using the categories 0, 1, 2 and 3 or more. **(3)**

(b) Calculate the test statistic and the degrees of freedom, and state the conclusion at the 5% level of significance, given that the critical value is 5.991. **(4)**

(Total for Question 8 is 7 marks)

9 The random variable X ~ B(12, p). A test of H0: p = 0.25 against H1: p > 0.25 is carried out at the 10% significance level.

(a) Find the critical region for the test. **(2)**

(b) State the probability of a Type I error, to 3 significant figures. **(2)**

(c) Find the probability of a Type II error when the true value of p is 0.7, to 3 significant figures. **(2)**

(Total for Question 9 is 6 marks)

10 The number of faults in a length of fabric is modelled by a Poisson distribution with mean λ. A test of H0: $\lambda = 6$ against H1: $\lambda < 6$ is carried out at the 5% significance level, using a single observation X of the number of events.

(a) Find the critical region for the test. **(3)**

(b) State the probability of a Type I error for this test, to 3 significant figures. **(2)**

(c) Given that the true value of λ is 3, find the probability of a Type II error, to 3 significant figures. **(2)**

(Total for Question 10 is 7 marks)

11 A market researcher classifies the answers to a survey question into 4 categories. In a random sample of 120 observations the frequencies in the categories are 24, 36, 30, 30. A model predicts that the categories occur in the ratio 1:1:1:1. A test at the 5% significance level is to be carried out to see whether the model fits.

(a) State the hypotheses for the test and find the expected frequency for each category. **(3)**

(b) Calculate the test statistic, to 3 significant figures, and state the number of degrees of freedom. **(3)**

(c) The critical value of chi^2 for 3 degrees of freedom at the 5% level is 7.815. State your conclusion in context. **(3)**

(Total for Question 11 is 9 marks)

12 A geneticist classifies the offspring of an experiment into 4 categories. In a random sample of 67 observations the frequencies in the categories are 22, 13, 16, 16. A model predicts that the categories occur in the ratio 1:1:1:1. A test at the 5% significance level is to be carried out to see whether the model fits.

(a) State the hypotheses for the test and find the expected frequency for each category. **(3)**

(b) Calculate the test statistic, to 3 significant figures, and state the number of degrees of freedom. **(3)**

(c) The critical value of chi^2 for 3 degrees of freedom at the 5% level is 7.815. State your conclusion in context. **(3)**

(Total for Question 12 is 9 marks)

13 A teacher classifies the results of a class test into 4 categories. In a random sample of 162 observations the frequencies in the categories are 46, 48, 32, 36. A model predicts that the categories occur in the ratio 1:1:1:1. A test at the 1% significance level is to be carried out to see whether the model fits.

(a) State the hypotheses for the test and find the expected frequency for each category. **(3)**

(b) Calculate the test statistic, to 3 significant figures, and state the number of degrees of freedom. **(3)**

(c) The critical value of chi^2 for 3 degrees of freedom at the 1% level is 11.345. State your conclusion in context. **(3)**

(Total for Question 13 is 9 marks)

14 A fair coin-like device is claimed to land on 'success' with probability 0.15. A tester believes the probability of success is actually lower. Let X be the number of trials needed to get the first success, so under the claim X ~ Geo(0.15). The first success occurs on trial 24.

(a) State suitable hypotheses for the test, in terms of p. **(2)**

(b) Find $P(X \geq 24)$ assuming H0 is true, and test the tester's belief at the 10% level. **(3)**

(c) Find the critical region for the test at the 10% level, and the probability of a Type II error if the true value of p is 0.08, to 3 significant figures. **(4)**

(Total for Question 14 is 9 marks)

15 A researcher records the results of 147 people classified by group and by outcome, as shown. Rows: Morning class: Pass 25, Merit 29, Distinction 42; Afternoon class: Pass 15, Merit 17, Distinction 19. A test is carried out at the 1% level to see whether the outcome is independent of the group.

	Pass	Merit	Distinction
Morning class	25	29	42
Afternoon class	15	17	19

(a) State the hypotheses, and calculate the expected frequency for the first cell in each row. **(3)**

(b) Find the value of the test statistic, to 3 significant figures, and the number of degrees of freedom. **(3)**

(c) Given that the critical value is 9.21, state the conclusion of the test in context. **(3)**

(Total for Question 15 is 9 marks)

16 A researcher records the results of 174 people classified by group and by outcome, as shown. Rows: Urban: Walk 31, Bus 30, Car 39; Rural: Walk 29, Bus 24, Car 21. A test is carried out at the 1% level to see whether the outcome is independent of the group.

	Walk	Bus	Car
Urban	31	30	39
Rural	29	24	21

(a) State the hypotheses, and calculate the expected frequency for the first cell in each row. **(3)**

(b) Find the value of the test statistic, to 3 significant figures, and the number of degrees of freedom. **(3)**

(c) Given that the critical value is 9.21, state the conclusion of the test in context. **(3)**

(Total for Question 16 is 9 marks)

17 A researcher records the results of 144 people classified by group and by outcome, as shown. Rows: Morning class: Pass 29, Merit 23, Distinction 34; Afternoon class: Pass 19, Merit 15, Distinction 24. A test is carried out at the 5% level to see whether the outcome is independent of the group.

	Pass	Merit	Distinction
Morning class	29	23	34
Afternoon class	19	15	24

(a) State the hypotheses, and calculate the expected frequency for the first cell in each row. **(3)**

(b) Find the value of the test statistic, to 3 significant figures, and the number of degrees of freedom. **(3)**

(c) Given that the critical value is 5.991, state the conclusion of the test in context. **(3)**

(Total for Question 17 is 9 marks)

18 A manufacturer states that the weight, X grams, of bags of flour it produces can be modelled by $X \sim N(1000, 15^2)$. A quality inspector weighs a random sample of 200 bags and classifies each into one of five weight classes, obtaining the observed frequencies below.

Weight class: under 985g, 985g up to 995g, 995g up to 1005g, 1005g up to 1015g, 1015g or more.
Observed frequency: 40, 38, 50, 45, 27.

(a) Show that the z-value corresponding to a weight of 985g is -1, and find the z-values corresponding to 995g, 1005g and 1015g. **(2)**

(b) Hence show that $P(X < 985) = 0.1587$, and find the probability that a bag falls into each of the other four weight classes, given that $\Phi(0.3333) = 0.6306$. **(3)**

(c) Calculate the expected frequency for each weight class, to 2 decimal places. **(2)**

(d) Calculate the value of the chi-squared test statistic. **(3)**

(e) Using a 5% significance level, state the number of degrees of freedom, the critical value, and your conclusion in context. **(2)**

(f) The manufacturer's engineer suspects that, although the mean weight is indeed 1000g, the standard deviation may not be exactly 15g and should instead be estimated from the sample. State, with a reason, how the number of degrees of freedom for the test would change. **(2)**

(Total for Question 18 is 14 marks)

19 A tutoring company offers one-to-one sessions on four afternoons each week: Monday, Tuesday, Wednesday and Thursday. It records the day chosen by a random sample of 120 GCSE students and 60 A-level students, to test whether preferred day is associated with key stage. The results are given below.

GCSE: Monday 38, Tuesday 42, Wednesday 32, Thursday 8, row total 120.
A-level: Monday 22, Tuesday 19, Wednesday 15, Thursday 4, row total 60.
Column totals: Monday 60, Tuesday 61, Wednesday 47, Thursday 12, grand total 180.

(a) Calculate the expected frequency for each of the 8 cells of the original table, showing your method for at least two cells. **(3)**

(b) Identify which cell has an expected frequency below 5, and explain why, and with which neighbouring column, it should be merged before the test statistic is calculated. **(2)**

(c) Complete the table of observed and expected frequencies for the merged 2 by 3 table, with columns Monday, Tuesday and 'Wednesday or Thursday'. **(2)**

(d) State the number of degrees of freedom for the test after merging. **(1)**

(e) Calculate the value of the chi-squared test statistic for the merged table. **(3)**

(f) Using the critical value chi-squared$_2$(0.05) = 5.991, test at the 5% significance level and state your conclusion in context. **(2)**

(Total for Question 19 is 13 marks)

20 A sample of 120 pet owners is classified by pet type (dog or cat) and whether they hold pet insurance, to test for association between the two.

Dog owners: insured 42, not insured 18, row total 60.
Cat owners: insured 24, not insured 36, row total 60.
Column totals: insured 66, not insured 54, grand total 120.

For a chi-squared test, the standardised residual for a cell is defined as $e = (O-E)/\sqrt{E}$. Unlike the term $(O-E)^2/E$ used in the test statistic itself, the sign of e shows whether that cell's observed frequency was higher or lower than expected.

(a) Calculate the expected frequency for each cell, showing your method for one cell. **(1)**

(b) Calculate the standardised residual, $e = (O-E)/\sqrt{E}$, for each cell, giving each to 2 decimal places. **(3)**

(c) Verify that summing the squares of your four standardised residuals gives the same chi-squared test statistic as the standard formula, and state its value. Comment on the significance of this value at the 5%, 1% and 0.5% levels, given critical values of 3.841, 6.635 and 7.879 respectively (df = 1). **(2)**

(d) Using the signs of the standardised residuals, describe the nature of the association between pet type and having pet insurance. **(1)**

(Total for Question 20 is 7 marks)

Mark scheme · 2.6 Further Statistics: Hypothesis Testing and Chi-Squared Depth

Question 1

(a) H0 for a test of association states precisely that the two categorical variables are independent (oe) **B1**

(a) i.e. for every row category and column category, P(that row and that column) = P(that row) * P(that column) **B1**

(a) Answer: H0 is exactly the statement that the two variables are independent: P(row and column) = P(row) * P(column) for every cell.

(b) under independence, P(row 1 and column 1) = (R1/n) * (C1/n) **M1**

(b) cso: expected frequency = n * P(row 1 and column 1) = n * (R1/n) * (C1/n) = R1*C1/n, as required **A1**

(b) Answer: E(row 1, col 1) = n * (R1/n) * (C1/n) = R1*C1/n, matching the standard formula.

(c) a test statistic of 0 means every observed frequency in the sample exactly equals its expected (independence) value, so this sample gives the strongest possible evidence in favour of H0 **B1**

(c) however this only describes the particular sample collected; it does not prove that the underlying population variables are truly independent, since a different sample could show some discrepancy. It is evidence for, not proof of, independence **B1**

(c) Answer: Not fully accurate: a value of 0 gives the strongest possible sample evidence for independence, but it does not prove the population variables are independent.

Question 2

(a) when a parameter (e.g. a Poisson mean or a binomial p) is estimated from the sample, it is deliberately chosen (e.g. by matching the sample mean) so that the fitted distribution reproduces key features of that very sample as closely as possible **B1**

(a) this means the resulting expected frequencies tend to be closer to the observed frequencies, and the test statistic tends to be smaller, than if a genuinely independent, pre-specified parameter value had been used **B1**

(a) reducing the degrees of freedom by 1 for each estimated parameter compensates for this artificial closeness of fit, so the test keeps its stated significance level **B1**

(a) Answer: Estimating a parameter from the data makes the model fit that data unusually well by construction; losing one degree of freedom per estimated parameter corrects for this so the test keeps its stated significance level.

(b) (i) df = 6 - 1 = 5 **B1**

(b) (ii) df = 6 - 1 - 1 = 4 **B1**

(b) (iii) df = 6 - 1 - 2 = 3 **B1**

(b) Answer: (i) 5 (ii) 4 (iii) 3

Question 3

(a) H0: attending the Saturday club and passing the mock exam are independent (no association) **B1**

(a) H1: attending the Saturday club and passing the mock exam are associated **B1**

(a) Answer: H0: club attendance and passing the mock are independent. H1: they are associated.

(b) uses E = (row total * column total) / grand total **M1**

(b) all four correct: 48, 32, 42, 28 **A1**

(b) Answer: 48, 32, 42, 28

(c) uses X^2 = sum of $(O-E)^2/E$ **M1**

(c) at least 3 terms correct (unsimplified): 144/48, 144/32, 144/42, 144/28 **A1**

(c) cao, awrt 16.07 **A1**

(c) Answer: X^2 = awrt 16.07 (accept 16.1)

(d) df = (2-1)(2-1) = 1; 16.07 > 3.841, so reject H0: there is significant evidence of an association between attending the Saturday club and passing the mock exam **A1**

(d) since 16.07 also exceeds 7.879 (the 0.5% critical value), this is very strong evidence, significant even at the 0.5% level **A1**

(d) Answer: df = 1; 16.07 exceeds even the 0.5% critical value, so there is very strong evidence of an association.

(e) Yates' correction is usually recommended when the sample size (or one or more expected frequencies) is small, since the discrete chi-squared statistic is then poorly approximated by the continuous chi-squared distribution **B1**

(e) the correction can only ever reduce the test statistic slightly (each |O-E| is reduced by 0.5 before squaring); since 16.07 so greatly exceeds even the most stringent critical value quoted (7.879), a small reduction could not change the conclusion **B1**

(e) Answer: Recommended when expected frequencies/sample size are small; here the margin over the critical value is so large that a small correction cannot change the outcome.

(f) the correction compensates for approximating a discrete test statistic by the continuous chi-squared distribution, and this approximation is least accurate specifically when df = 1 **B1**

(f) for tables with more categories (df > 1) the approximation is already good enough without correction, and there is no single agreed way to extend an analogous correction across several cells, so it is conventionally omitted **B1**

(f) Answer: The df = 1 approximation is the one that needs correcting; for larger tables the chi-squared approximation is already adequate and no standard correction is defined.

Question 4

(a) uses X^2 = sum of $(O-E)^2/E$ with E = 12.5 in every category **M1**

(a) cao: $X^2 = 4.72$ **A1**

(a) 4.72 < 7.815, so do not reject H0: no significant evidence against the categories being equally likely **A1**

(a) Answer: $X^2 = 4.72$; not significant at 5%.

(b) substitutes O' = 10*O and E' = 10*E into the formula: sum of $(10*O - 10*E)^2 / (10*E)$ **M1**

(b) simplifies: sum of $100*(O-E)^2 / (10*E) = 10 *$ sum of $(O-E)^2/E$ **M1**

(b) cso: new test statistic = 10 * 4.72 = 47.2 **A1**

(b) Answer: New test statistic = 10 * 4.72 = 47.2

(c) 47.2 > 7.815, so the result is now significant at the 5% level, in contrast to part (a) **A1**

(c) the exact same relative (proportional) pattern of deviation from the expected frequencies, which was not significant for n = 50, becomes highly significant purely because the sample size increased **B1**

(c) this shows that chi-squared significance reflects how confidently a deviation can be detected, not the size or practical importance of the deviation itself; a very large sample can make even a small, practically unimportant deviation 'statistically significant' **B1**

(c) Answer: 47.2 > 7.815, now significant; the same proportional pattern becomes significant purely because of a larger sample size, showing significance is not the same as practical importance.

Question 5

(a) 8.24 > 5.991, so reject H0 **A1**

(a) there is significant evidence of an association between the area a resident lives in and their preferred collection frequency **A1**

(a) Answer: 8.24 > 5.991, reject H0; significant evidence of an association between area and preferred collection frequency.

(b) rejecting H0 when H0 is actually true (oe) **B1**

(b) in context: concluding that there is an association between area and preferred collection frequency when, in reality, no such association exists **B1**

(b) Answer: Concluding there is an association between area and collection preference when in reality there is none.

(c) 0.05 **B1**

(c) Answer: 0.05

(d) lower the significance level (e.g. use 1% instead of 5%) **B1**

(d) this would increase the probability of a Type II error, since it becomes harder to reject H0, making it more likely that a real association is missed (not detected) **B1**

(d) Answer: Use a lower significance level (e.g. 1%); this increases the probability of a Type II error.

(e) a larger sample gives more precise, reliable estimates of the expected frequencies and increases the power of the test **B1**

(e) this makes the test more able to detect a genuine association where one exists, reducing the chance of a false 'no association' conclusion (a Type II error) **B1**

(e) Answer: Larger samples increase the power of the test, making a genuine association more likely to be detected, so Type II error probability falls.

(f) all expected frequencies (under H0, for each area/collection-frequency combination) must be at least 5 **B1**

(f) Answer: All expected frequencies for the area/collection-frequency cells must be at least 5.

Question 6

(a) $P(X = j) = e^{-1.504}\ 1.504^j/j!$ for j = 0, 1, 2 **M1**

(a) $P(X \geq 3) = 1 - P(X \leq 2)$ **M1**

(a) expected: 26.45, 39.78, 29.91, 22.87 **A1**

(a) Answer: 26.45, 39.78, 29.91, 22.87

(b) chi^2 = sum of $(O - E)^2/E$ **M1**

(b) 0.995 **A1**

(b) degrees of freedom = 4 - 1 - 1 = 2 (the mean was estimated from the data) **B1**

(b) do not reject H0: the Poisson model is a suitable fit **A1**

(b) Answer: chi^2 = 0.995, nu = 2, do not reject H0

Question 7

(a) H0: $p = 0.25$ **B1**

(a) H1: $p < 0.25$ **B1**

(a) Answer: H0: $p = 0.25$, H1: $p < 0.25$

(b) $P(X \geq 13) = (1 - 0.25)^{13 - 1}$ **M1**

(b) 0.0317 **A1**

(b) reject H0 **A1**

(b) Answer: 0.0317; reject H0

(c) finds the smallest d with $(1 - 0.25)^{d - 1} \leq 0.05$ **M1**

(c) critical region $X \geq 12$ **A1**

(c) $P(X \leq 11 \mid p = 0.1) = 1 - (1 - 0.1)^{11}$ **M1**

(c) 0.686 **A1**

(c) Answer: $X \geq 12$; Type II error 0.686

Question 8

(a) $P(X = j) = e^{-0.946}\ 0.946^j/j!$ for $j = 0, 1, 2$ **M1**

(a) $P(X \geq 3) = 1 - P(X \leq 2)$ **M1**

(a) expected: 57.86, 54.73, 25.89, 10.53 **A1**

(a) Answer: 57.86, 54.73, 25.89, 10.53

(b) chi^2 = sum of $(O - E)^2/E$ **M1**

(b) 0.944 **A1**

(b) degrees of freedom = 4 - 1 - 1 = 2 (the mean was estimated from the data) **B1**

(b) do not reject H0: the Poisson model is a suitable fit **A1**

(b) Answer: $chi^2 = 0.944$, nu = 2, do not reject H0

Question 9

(a) uses B(12, 0.25) cumulative probabilities **M1**

(a) critical region $X \geq 6$ **A1**

(a) Answer: critical region $X \geq 6$

(b) P(X in critical region | p = 0.25) **M1**

(b) 0.0544 **A1**

(b) Answer: 0.0544

(c) P(X not in critical region | p = 0.7) **M1**

(c) 0.0386 **A1**

(c) Answer: 0.0386

Question 10

(a) uses Po(6) probabilities in the appropriate tail **M1**

(a) $P(X \leq 1) = 0.0174 \leq 0.05$ **A1**

(a) critical region $X \leq 1$ **A1**

(a) Answer: critical region $X \leq 1$

(b) the probability of a Type I error = P(X in the critical region | H0 true) **M1**

(b) 0.0174 **A1**

(b) Answer: 0.0174

(c) P(X not in the critical region | $\lambda = 3$) **M1**

(c) 0.801 **A1**

(c) Answer: 0.801

Question 11

(a) H0: the model (ratio 1:1:1:1) is a suitable fit; H1: it is not **B1**

(a) expected = N x (ratio part)/4 **M1**

(a) expected: 30, 30, 30, 30 **A1**

(a) Answer: 30, 30, 30, 30

(b) sum of $(O - E)^2/E$ **M1**

(b) $chi^2 = 2.40$ **A1**

(b) degrees of freedom = 4 - 1 = 3 **B1**

(b) Answer: $chi^2 = 2.40$, nu = 3

(c) compares 2.40 with 7.815 **M1**

(c) do not reject H0 **A1**

(c) there is no evidence against the model: the data are consistent with the ratio **B1**

(c) Answer: do not reject H0

Question 12

(a) H0: the model (ratio 1:1:1:1) is a suitable fit; H1: it is not **B1**

(a) expected = N x (ratio part)/4 **M1**

(a) expected: 16.75, 16.75, 16.75, 16.75 **A1**

(a) Answer: 16.75, 16.75, 16.75, 16.75

(b) sum of $(O - E)^2/E$ **M1**

(b) $chi^2 = 2.55$ **A1**

(b) degrees of freedom = 4 - 1 = 3 **B1**

(b) Answer: $chi^2 = 2.55$, nu = 3

(c) compares 2.55 with 7.815 **M1**

(c) do not reject H0 **A1**

(c) there is no evidence against the model: the data are consistent with the ratio **B1**

(c) Answer: do not reject H0

Question 13

(a) H0: the model (ratio 1:1:1:1) is a suitable fit; H1: it is not **B1**

(a) expected = N x (ratio part)/4 **M1**

(a) expected: 40.5, 40.5, 40.5, 40.5 **A1**

(a) Answer: 40.5, 40.5, 40.5, 40.5

(b) sum of $(O - E)^2/E$ **M1**

(b) $chi^2 = 4.42$ **A1**

(b) degrees of freedom = 4 - 1 = 3 **B1**

(b) Answer: $chi^2 = 4.42$, nu = 3

(c) compares 4.42 with 11.345 **M1**

(c) do not reject H0 **A1**

(c) there is no evidence against the model: the data are consistent with the ratio **B1**

(c) Answer: do not reject H0

Question 14

(a) H0: p = 0.15 **B1**

(a) H1: p < 0.15 **B1**

(a) Answer: H0: p = 0.15, H1: p < 0.15

(b) $P(X \geq 24) = (1 - 0.15)^{24 - 1}$ **M1**

(b) 0.0238 **A1**

(b) reject H0 **A1**

(b) Answer: 0.0238; reject H0

(c) finds the smallest d with $(1 - 0.15)^{d - 1} \leq 0.1$ **M1**

(c) critical region $X \geq 16$ **A1**

(c) $P(X \leq 15 \mid p = 0.08) = 1 - (1 - 0.08)^{15}$ **M1**

(c) 0.714 **A1**

(c) Answer: X ≥ 16; Type II error 0.714

Question 15

(a) H0: outcome and group are independent; H1: they are associated **B1**

(a) expected = (row total x column total)/grand total **M1**

(a) $E_{11} = 1280/49 = 26.12$, $E_{21} = 680/49 = 13.88$ **A1**

(a) Answer: $E_{11} = 26.12$, $E_{21} = 13.88$

(b) sum of $(O - E)^2/E$ over all six cells **M1**

(b) $chi^2 = 0.582$ **A1**

(b) degrees of freedom = (2 - 1)(3 - 1) = 2 **B1**

(b) Answer: $chi^2 = 0.582$, nu = 2

(c) compares 0.582 with 9.21 **M1**

(c) do not reject H0 **A1**

(c) there is no evidence of an association between group and outcome **B1**

(c) Answer: do not reject H0

Question 16

(a) H0: outcome and group are independent; H1: they are associated **B1**

(a) expected = (row total x column total)/grand total **M1**

(a) $E_{11} = 1000/29 = 34.48$, $E_{21} = 740/29 = 25.52$ **A1**

(a) Answer: $E_{11} = 34.48$, $E_{21} = 25.52$

(b) sum of $(O - E)^2/E$ over all six cells **M1**

(b) $chi^2 = 2.30$ **A1**

(b) degrees of freedom = (2 - 1)(3 - 1) = 2 **B1**

(b) Answer: $chi^2 = 2.30$, nu = 2

(c) compares 2.30 with 9.21 **M1**

(c) do not reject H0 **A1**

(c) there is no evidence of an association between group and outcome **B1**

(c) Answer: do not reject H0

Question 17

(a) H0: outcome and group are independent; H1: they are associated **B1**

(a) expected = (row total x column total)/grand total **M1**

(a) $E_{11} = 86/3 = 28.67$, $E_{21} = 58/3 = 19.33$ **A1**

(a) Answer: $E_{11} = 28.67$, $E_{21} = 19.33$

(b) sum of $(O - E)^2/E$ over all six cells **M1**

(b) $chi^2 = 0.0491$ **A1**

(b) degrees of freedom = (2 - 1)(3 - 1) = 2 **B1**

(b) Answer: $chi^2 = 0.0491$, nu = 2

(c) compares 0.0491 with 5.991 **M1**

(c) do not reject H0 **A1**

(c) there is no evidence of an association between group and outcome **B1**

(c) Answer: do not reject H0

Question 18

(a) z = (985-1000)/15 = -1 shown **M1**

(a) other z-values correct: -1/3 (awrt -0.33), 1/3 (awrt 0.33), 1 **A1**

(a) Answer: z = -1, -0.333, 0.333, 1

(b) P(X<985) = Phi(-1) = 1 - Phi(1) = 1 - 0.8413 = 0.1587, shown **M1**

(b) uses Phi(-0.3333) = 1 - 0.6306 = 0.3694 and subtraction to find the middle classes **M1**

(b) all five probabilities correct: 0.1587, 0.2107, 0.2612, 0.2107, 0.1587 **A1**

(b) Answer: 0.1587, 0.2107, 0.2612, 0.2107, 0.1587

(c) multiplies each probability by 200 **M1**

(c) all correct: 31.74, 42.14, 52.24, 42.14, 31.74 **A1**

(c) Answer: 31.74, 42.14, 52.24, 42.14, 31.74 (sum = 200.00)

(d) uses X^2 = sum of $(O-E)^2/E$ with the correct O and E values **M1**

(d) at least 3 terms correct **A1**

(d) awrt 3.55 (allow awrt 3.56 if unrounded calculator probabilities are used) **A1**

(d) Answer: X^2 = awrt 3.55 (3.56 using unrounded expected frequencies)

(e) df = 5 - 1 = 4 (both mean and standard deviation were fixed, claimed values, not estimated); critical value chi-squared$_4$(0.05) = 9.488 **B1**

(e) ft: 3.55 < 9.488, so do not reject H0; there is no significant evidence against the manufacturer's claimed $N(1000, 15^2)$ model for the bag weights **A1**

(e) Answer: df = 4, critical value 9.488; do not reject H0, the claimed Normal model is a good fit.

(f) df would become 5 - 1 - 1 = 3 **B1**

(f) because one additional parameter (the standard deviation) would then be estimated from the sample data, and each estimated parameter reduces the degrees of freedom by a further 1 **B1**

(f) Answer: df would decrease to 3, since σ would then be estimated from the data.

Question 19

(a) uses E = (row total * column total) / grand total **M1**

(a) at least 4 correct **A1**

(a) all 8 correct: 40, 40.67, 31.33, 8, 20, 20.33, 15.67, 4 **A1**

(a) Answer: GCSE: 40, 40.67, 31.33, 8. A-level: 20, 20.33, 15.67, 4.

(b) the A-level, Thursday cell has expected frequency 4, which is below 5, so the validity condition for the chi-squared approximation is not met **B1**

(b) the whole Thursday column should therefore be merged with the neighbouring Wednesday column, giving a combined 'Wednesday or Thursday' category for both rows **B1**

(b) Answer: A-level/Thursday has E = 4 < 5; merge the Thursday column into Wednesday.

(c) combines the observed frequencies: GCSE 38, 42, 40; A-level 22, 19, 19 **M1**

(c) combines the expected frequencies correctly: GCSE 40, 40.67, 39.33; A-level 20, 20.33, 19.67 **A1**

(c) Answer: O: GCSE 38, 42, 40; A-level 22, 19, 19. E: GCSE 40, 40.67, 39.33; A-level 20, 20.33, 19.67.

(d) df = (2-1)(3-1) = 2 **B1**

(d) Answer: df = 2

(e) uses X^2 = sum of $(O-E)^2/E$ on the merged table **M1**

(e) at least 3 terms correct **A1**

(e) cao, awrt 0.465 **A1**

(e) Answer: X^2 = awrt 0.465

(f) 0.465 < 5.991 **A1**

(f) ft: do not reject H0; no significant evidence of an association between key stage and preferred session day **A1**

(f) Answer: 0.465 < 5.991, do not reject H0; no evidence of association.

Question 20

(a) E = 33, 27, 33, 27 (e.g. 60*66/120 = 33) **B1**

(a) Answer: 33, 27, 33, 27

(b) uses $e = (O-E)/\sqrt{E}$ for at least one cell **M1**

(b) at least 2 correct **A1**

(b) all four correct: 1.57, -1.73, -1.57, 1.73 **A1**

(b) Answer: Dog/insured: 1.57. Dog/not insured: -1.73. Cat/insured: -1.57. Cat/not insured: 1.73.

(c) sums the squares of the four residuals: $1.57^2 + 1.73^2 + 1.57^2 + 1.73^2$ **M1**

(c) awrt 10.9, which exceeds all three critical values, so the result is significant even at the 0.5% level **A1**

(c) Answer: X^2 = awrt 10.9, significant at the 5%, 1% and 0.5% levels.

(d) dog owners are more likely than expected under independence to have insurance (positive residual) and less likely than expected to be uninsured, while cat owners show the opposite pattern; i.e. dog owners are more likely to be insured than cat owners **B1**

(d) Answer: Dog owners are more likely to have pet insurance than cat owners.