2.3 Further Statistics: Chi-Squared Tests (Goodness of Fit and Contingency Tables): 30-minute Pack

Calculator allowed · about 30 minutes

Total Marks

Name: _____ Date: ____ / ____ / _____

Answer ALL questions. Show all your working.

Key facts: chi-squared tests

Original content written for Revision Library.

A chi-squared test compares observed frequencies O with the expected frequencies E predicted by a proposed model, using the test statistic X^2 = sum of (O - E)^2 / E, calculated over all classes (or cells) after any pooling. For a goodness-of-fit test to a fully specified distribution (for example a uniform distribution, or given proportions), degrees of freedom = (number of classes) - 1. If one or more parameters of the model (for example p for a binomial, or the mean for a Poisson) are estimated from the same sample used in the test, one further degree of freedom is lost for each parameter estimated. For an r by c contingency table testing independence between two variables, expected frequencies are E = (row total * column total) / grand total, and degrees of freedom = (r - 1)(c - 1). Classes with an expected frequency below 5 should be pooled (combined with a neighbouring class) before the test statistic is calculated, reducing the number of classes, and hence the degrees of freedom, accordingly. For any test with 1 degree of freedom, including every 2 by 2 contingency table, Yates' continuity correction is applied: X^2 = sum of (|O - E| - 0.5)^2 / E. The calculated test statistic is compared with the critical value from chi-squared tables at the chosen significance level and the stated degrees of freedom; if the test statistic exceeds the critical value, the null hypothesis is rejected.

1 A researcher records the results of 190 people classified by group and by outcome, as shown. Rows: Group X: Grade A 44, Grade B 23, Grade C 43; Group Y: Grade A 26, Grade B 14, Grade C 40. A test is carried out at the 5% level to see whether the outcome is independent of the group.

	Grade A	Grade B	Grade C
Group X	44	23	43
Group Y	26	14	40

(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 1 is 9 marks)

© Revision Library | revisionlibrary.co.uk

2 A researcher counts the number of events per interval over 157 intervals: 0 events 46 times, 1 event 57 times, 2 events 35 times, and 3 or more events 19 times. It is suggested that the number of events per interval follows a Poisson distribution. Taking the mean of the data to be 1.172 (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 2 is 7 marks)

---

3 A bakery in York claims that its daily sales split across four product types, cakes, pastries, bread and biscuits, in the ratio 4:3:2:1. On a randomly chosen day, 200 items are sold, with the following breakdown.

Cakes: 70. Pastries: 64. Bread: 44. Biscuits: 22.

(a) State suitable null and alternative hypotheses to test the bakery's claim. **(2)**

(b) Find the expected number of each product type sold, based on the claimed ratio. **(2)**

(c) Calculate the chi-squared test statistic. **(2)**

(d) State the number of degrees of freedom, explaining your answer, and the critical value at the 5% significance level. Hence test the bakery's claim, stating your conclusion clearly, in context. **(3)**

(Total for Question 3 is 9 marks)

© Revision Library | revisionlibrary.co.uk

Mark scheme · 2.3 Further Statistics: Chi-Squared Tests (Goodness of Fit and Contingency Tables): 30-minute Pack

Question 1

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

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

(a) $E_{11} = 770/19 = 40.53$, $E_{21} = 560/19 = 29.47$ **A1**

(a) Answer: $E_{11} = 40.53$, $E_{21} = 29.47$

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

(b) $chi^2 = 2.25$ **A1**

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

(b) Answer: $chi^2 = 2.25$, nu = 2

(c) compares 2.25 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 2

(a) $P(X = j) = e^{-1.172}\ 1.172^j/j!$ for j = 0, 1, 2 **M1**

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

(a) expected: 48.63, 56.99, 33.4, 17.98 **A1**

(a) Answer: 48.63, 56.99, 33.4, 17.98

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

(b) 0.277 **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: $chi^2 = 0.277$, nu = 2, do not reject H0

Question 3

(a) H0: sales occur in the ratio 4:3:2:1, as claimed **B1**

(a) H1: sales do not occur in the ratio 4:3:2:1 **B1**

(a) Answer: H0: sales occur in the ratio 4:3:2:1. H1: they do not.

(b) converts the ratio 4:3:2:1 to proportions 0.4, 0.3, 0.2, 0.1 and multiplies by 200 **M1**

(b) all four correct: 80, 60, 40, 20 **A1**

(b) Answer: Expected frequencies: 80, 60, 40, 20.

(c) attempts sum of $(O-E)^2/E$ across all 4 classes **M1**

(c) cao, awrt 2.12 **A1**

(c) Answer: Test statistic = awrt 2.12

(d) the ratio was fully specified (not estimated from the sample), so df = 4 - 1 = 3 **B1**

(d) critical value = 7.815 **B1**

(d) valid conclusion ft: since 2.12 < 7.815, do not reject H0; the data are consistent with the claimed ratio **B1**

(d) Answer: df = 3; critical value = 7.815; 2.12 < 7.815 so do not reject H0.

© Revision Library | revisionlibrary.co.uk