2.6 Further Statistics: Hypothesis Testing and Chi-Squared Depth: 30-minute Pack

EDEXCEL 9FM0 · Calculator allowed · about 30 minutes

Total Marks

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.

© Revision Library | revisionlibrary.co.uk

1 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 1 is 6 marks)

© Revision Library | revisionlibrary.co.uk

2 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 2 is 13 marks)

© Revision Library | revisionlibrary.co.uk

3 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 3 is 9 marks)

© Revision Library | revisionlibrary.co.uk

Mark scheme · 2.6 Further Statistics: Hypothesis Testing and Chi-Squared Depth: 30-minute Pack

Question 1

(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 2

(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.

© Revision Library | revisionlibrary.co.uk

Question 3

(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.

© Revision Library | revisionlibrary.co.uk