2.6 Further Statistics: Hypothesis Testing and Chi-Squared Depth: 1-hour Pack

EDEXCEL 9FM0 · Calculator allowed · about 60 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 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 1 is 10 marks)

---

2 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 2 is 7 marks)

© Revision Library | revisionlibrary.co.uk

3 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 ≥ 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 3 is 9 marks)

---

4 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 4 is 14 marks)

© Revision Library | revisionlibrary.co.uk

5 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 5 is 13 marks)

© Revision Library | revisionlibrary.co.uk

Mark scheme · 2.6 Further Statistics: Hypothesis Testing and Chi-Squared Depth: 1-hour Pack

Question 1

(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 2

(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 3

(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 ≥ 12; Type II error 0.686

Question 4

(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 $chi\text{-}squared_4(0.05)$ = 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.

© Revision Library | revisionlibrary.co.uk

Question 5

(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.

© Revision Library | revisionlibrary.co.uk