IB Math AI SL — Topic 4

Hypothesis Testing

Hypotheses & Significance · χ^2 Independence & Goodness of Fit · The t-test

Difficulty	EASY	MEDIUM	HARD
Questions	10	10	10
Total Marks	30	40	50

Scan me for help

Name:______________________________ Date:_______________

Score:__________/120

Instructions: Answer all questions. Show all working. Give answers correct to 3 significant figures unless stated otherwise. A graphic display calculator is required throughout; support GDC answers with the values entered and the feature used.

Scan me for help

Key Formulae

- **Chi-squared test statistic:** $\chi^2_{\text{calc}} = \sum \dfrac{(f_o - f_e)^2}{f_e}$, where f_o = observed frequency, f_e = expected frequency.
- **Expected frequency (independence):** $f_e = \dfrac{(\text{row total}) \times (\text{column total})}{\text{grand total}}$
- **Degrees of freedom (independence):** $\nu = (r-1)(c-1)$, where r = number of rows, c = number of columns.
- **Degrees of freedom (goodness of fit):** $\nu = k - 1$, where k = number of categories.
- **Two-sample t-test statistic:** $t = \dfrac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \quad s_p^2 = \dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}$
- **Decision rule (p-value method):** Reject H_0 if p-value $< \alpha$; do not reject H_0 if p-value $\geq \alpha$.
- **Decision rule (critical value method):** Reject H_0 if $\chi^2_{\text{calc}} > \chi^2_{\text{crit}}$ or $|t_{\text{calc}}| > t_{\text{crit}}$.

GDC Tips

- **Chi-squared independence / goodness of fit:** `STAT` → `TESTS` → χ^2`-Test` (independence) or χ^2 `GOF-Test` (goodness of fit); enter observed frequencies in list `L1` and expected frequencies in `L2`; read off χ^2 and `p-value` directly.
- **Two-sample t-test:** `STAT` → `TESTS` → `2-SampTTest`; enter $\bar{x}_1, s_1, n_1, \bar{x}_2, s_2, n_2$; set `Pooled: Yes`; choose $\neq \mu_2$ (two-tailed) or $< \mu_2$ / $> \mu_2$ (one-tailed) based on context; read off `t` and `p`.
- **Entering a contingency table for χ^2:** `2nd` → `MATRIX` → `EDIT` → `[A]`; set dimensions $r \times c$; enter observed counts row by row; GDC computes expected matrix `[B]` automatically after the test.
- **Reading the expected matrix:** After running χ^2`-Test`, press `2nd` → `MATRIX` → `[B]` → `ENTER` to display all expected frequencies for verification.
- **One- vs two-tailed selection:** If the research question uses "greater", "increased", or "decreased", select the matching one-tailed alternative; if it uses "different", "changed", or "differs", always select two-tailed ($\neq$).
- **Storing unrounded values:** After GDC output, store the unrounded p-value or t-statistic with `STO`→ before rounding, to avoid premature-rounding errors in later parts.

Common Mistakes to Avoid

1. **Wrong tail direction in the t-test.** Students automatically choose two-tailed regardless of context. Read the research question carefully: verbs such as "increased" or "decreased" signal a one-tailed test; "differs" or "changed" signal two-tailed. Selecting the wrong tail halves or doubles the p-value and changes the conclusion.

2. **Assertive conclusion language.** Writing "the variables are dependent" or "H_0 is proved false" loses the conclusion mark. Conclusions must be non-assertive: "there is sufficient evidence to suggest ..." or "there is insufficient evidence to suggest ...", matching the reject / do-not-reject decision.

3. **Comparison line omitted before the conclusion.** The R1 mark requires an explicit statement such as "$p = 0.0234 < 0.05 = \alpha$" or "$\chi^2_{\text{calc}} = 8.31 > \chi^2_{\text{crit}} = 5.99$" before any conclusion is drawn. Jumping straight to the conclusion earns R0 A0.

4. **Incorrect degrees of freedom for goodness of fit.** Students use $\nu = k$ instead of $\nu = k - 1$, or apply the independence formula $(r-1)(c-1)$ to a goodness-of-fit problem. For GOF with k categories, always use $\nu = k - 1$.

5. **Premature rounding of the expected frequency in the "show that" cell.** The expected value must be shown unrounded (e.g. $\frac{45\times 38}{120} = 14.25$) before stating it rounds to the printed target. Rounding intermediate values and then claiming the rounded figure matches the target is not accepted.

6. **Forgetting the pooled-variance assumption in the two-sample t-test.** The IB AI SL two-sample t-test assumes equal population variances (pooled). If the GDC is set to `Pooled: No`, the degrees of freedom and p-value differ and the answer will not match the mark scheme. Always set `Pooled: Yes` unless the question explicitly states otherwise.

Scan me for help

Section A — Easy

EASY

1. A researcher investigates whether the type of exercise preferred (running, swimming, cycling) is independent of age group (under 30, 30 and over) for gym members in Singapore. The observed frequencies are shown in the table below.

	Running	Swimming	Cycling
Under 30	28	15	17
30 and over	22	25	13

State the null hypothesis H_0 and alternative hypothesis H_1 for a chi-squared test of independence. **[3 marks]**

2. A chi-squared test of independence is carried out on a 3×4 contingency table.

(a) Write down the number of degrees of freedom. **[1 marks]**

(b) The chi-squared statistic is $\chi^2 = 14.3$ and the critical value at the 5% significance level is 12.592. Determine whether H_0 should be rejected. Justify your answer. **[2 marks]**

3. A factory produces coloured sweets in four colours: red, blue, green, and yellow. The manufacturer claims the sweets are produced in equal proportions. A sample of 120 sweets is collected and the observed frequencies are shown below.

Colour	Red	Blue	Green	Yellow
Observed	34	28	31	27

(a) Write down the expected frequency for each colour under the manufacturer's claim. **[1 marks]**

(b) Write down the number of degrees of freedom for a goodness of fit test. **[1 marks]**

(c) Find the p-value for the goodness of fit test. **[1 marks]**

4. A teacher believes that students who study using flashcards score higher on vocabulary tests than those who use textbooks. Two independent groups of students are tested. The teacher decides to use a t-test to compare the two group means.
State the null hypothesis H_0 and alternative hypothesis H_1 for this test, using μ_1 for the flashcard group mean and μ_2 for the textbook group mean. **[3 marks]**

5. A two-sample t-test is performed to compare the mean daily water consumption (in litres) of adults in two cities, City A and City B. The test is two-tailed at the 5% significance level.
The GDC gives a p-value of 0.0312.

(a) Write down the comparison needed to reach a conclusion. **[1 marks]**

(b) State the conclusion of the test in context. **[2 marks]**

6. A survey of 200 university students in Brazil records whether each student prefers tea or coffee, and whether they study in the morning or the evening. The results are shown in the table below.

	Tea	Coffee
Morning	54	66
Evening	36	44

Show that the expected frequency of morning tea drinkers is 45, correct to the nearest integer. **[3 marks]**

7. For a goodness of fit test on a data set with 6 categories, write down the number of degrees of freedom. **[1 marks]**
A goodness of fit test yields a p-value of 0.182 at the 5% significance level.
State the conclusion of the test. **[2 marks]**

8. Two independent random samples of plant heights (in cm) are collected from two different greenhouses, Greenhouse P and Greenhouse Q. A two-sample t-test is used to test whether the mean heights differ. The significance level is 5%.

	Greenhouse P	Greenhouse Q
Sample size	18	20
Sample mean (cm)	34.2	31.6
Sample standard deviation (cm)	4.1	3.8

Find the p-value for this test and state whether H_0 is rejected. **[3 marks]**

9. A chi-squared test of independence is carried out on the data in a 2×3 contingency table. The chi-squared statistic obtained is $\chi^2_{\text{calc}} = 3.84$. The critical value at the 10% significance level is $\chi^2_{\text{crit}} = 4.605$.

(a) Write down the degrees of freedom. **[1 marks]**

(b) Determine whether H_0 should be rejected at the 10% significance level. Justify your answer. **[2 marks]**

10. A nutritionist claims that the proportion of adults choosing each of three breakfast options (cereal, toast, eggs) follows the ratio $2 : 3 : 1$. In a sample of 120 adults, the observed frequencies are: cereal 42, toast 58, eggs 20.

(a) Calculate the expected frequency for each breakfast option under the nutritionist's claim. **[2 marks]**

(b) Write down the number of degrees of freedom for a goodness of fit test on these data. **[1 marks]**

Scan me for help

Section B — Medium

MEDIUM

11. A school counsellor surveys 120 students and records whether they prefer *morning* or *afternoon* study sessions and their year group (*Year 1*, *Year 2*, or *Year 3*). The results are shown in the table below.

	Year 1	Year 2	Year 3
Morning	22	18	30
Afternoon	18	22	10

A chi-squared test for independence is carried out at the 5% significance level.

(a) State the null and alternative hypotheses for this test. **[1 marks]**

(b) Show that the expected frequency for Year 3 / Morning is 22.5. **[1 marks]**

(c) Find the p-value for this test. **[1 marks]**

(d) State the conclusion of the test in context. **[1 marks]**

12. A botanist claims that seeds from a particular plant germinate in four colours in the ratio Red : Yellow : White : Purple $= 3 : 2 : 2 : 1$. She plants 160 seeds and observes the following counts.

Colour	Red	Yellow	White	Purple
Observed	64	38	35	23

A goodness-of-fit test is carried out at the 10% significance level.

(a) Write down the expected frequency for each colour. **[1 marks]**

(b) Write down the number of degrees of freedom. **[1 marks]**

(c) Find the p-value for this test and state the conclusion in context. **[2 marks]**

13. Two groups of athletes follow different hydration programmes. Their post-exercise heart rates (beats per minute) are recorded below.

Programme A	88	92	85	90	94	87
Programme B	95	99	101	97	93	96

A two-sample t-test is performed at the 5% significance level to test whether the mean heart rate differs between the two programmes.

(a) State the null and alternative hypotheses for this test. **[1 marks]**

(b) State one assumption required for this test to be valid. **[1 marks]**

(c) Find the p-value. **[1 marks]**

(d) State the conclusion in context. **[1 marks]**

14. A factory produces light bulbs. Quality control records state that bulb lifetimes are normally distributed with mean $\mu = 1200$ hours and standard deviation $\sigma = 80$ hours. A new manufacturing process is introduced and a sample of 10 bulbs gives the following lifetimes (hours):

$$1230,\ 1195,\ 1260,\ 1240,\ 1215,\ 1280,\ 1250,\ 1205,\ 1270,\ 1255$$

A one-sample t-test is carried out at the 5% significance level to test whether the mean lifetime has *increased*.

(a) State the null and alternative hypotheses. **[1 marks]**

(b) Find the p-value. **[1 marks]**

(c) Hence determine, with justification, whether H_0 should be rejected. **[2 marks]**

15. A researcher surveys 200 adults in two cities, Aldenvale and Briston, asking whether they support a new public transport policy (*Support* or *Oppose*). The observed counts are given below.

	Support	Oppose	Total
Aldenvale	74	46	120
Briston	38	42	80
Total	112	88	200

A chi-squared test for independence is performed at the 5% significance level.

(a) Show that the expected number of Briston residents who Support the policy is 44.8. **[1 marks]**

(b) Write down the degrees of freedom. **[1 marks]**

(c) Find the p-value and state the conclusion in context. **[2 marks]**

16. A six-sided die is rolled 120 times. A student believes the die is fair. The observed frequencies are shown below.

Face	1	2	3	4	5	6
Observed	24	16	22	18	26	14

A goodness-of-fit test is carried out at the 10% significance level.

(a) Write down the expected frequency for each face. **[1 marks]**

(b) Find the p-value for the test. **[1 marks]**

(c) Determine, with justification, the largest significance level at which the null hypothesis would **not** be rejected. **[2 marks]**

17. A nutritionist tests whether a new diet affects resting blood glucose levels (mmol/L). She measures 8 participants before and after the diet. The *differences* (before − after) are:

$$0.4,\ 0.7,\ -0.1,\ 0.5,\ 0.9,\ 0.3,\ 0.6,\ 0.2$$

She uses a one-sample t-test on the differences to test whether the diet *reduces* glucose levels at the 5% significance level.

(a) State H_0 and H_1 in terms of the population mean difference μ_d. **[1 marks]**

(b) Find the p-value for this test. **[1 marks]**

(c) Hence state the conclusion in context, justifying your answer. **[2 marks]**

18. A university lecturer records the number of times students access online lecture notes in one week. She believes access frequency follows the distribution below.

Accesses per week	0	1	2	3	4+
Expected proportion	0.10	0.20	0.35	0.25	0.10
Observed frequency	8	22	40	28	2

A goodness-of-fit test is performed at the 5% significance level on a sample of 100 students.

(a) Write down the expected frequencies. **[1 marks]**

(b) Write down the degrees of freedom. **[1 marks]**

(c) Find the p-value and state the conclusion in context. **[2 marks]**

19. Two independent groups of students complete a standardised reading test (score out of 100). Group X uses a digital reading app and Group Y uses traditional textbooks.

Group X (app): $72, 68, 75, 80, 65, 78, 71$

Group Y (books): $63, 70, 66, 58, 74, 61, 67$

A two-sample t-test is carried out at the 5% significance level to test whether the app *increases* the mean score.

(a) State the null and alternative hypotheses in terms of μ_X and μ_Y. **[1 marks]**

(b) Find the p-value. **[1 marks]**

(c) Hence state the conclusion in context, justifying your answer. **[2 marks]**

20. A travel agency surveys customers from three destinations (*Beach*, *Mountain*, *City*) and records whether they rated their trip *Excellent*, *Good*, or *Poor*. Partial results are shown below; the total number of customers surveyed is 180.

	Excellent	Good	Poor	Total
Beach	28	22	10	60
Mountain	20	25	15	60
City	24	18	18	60
Total	72	65	43	180

A chi-squared test for independence is carried out at the 5% significance level.

(a) Show that the expected frequency for the City / Poor cell is $\frac{43}{3} \approx 14.3$. **[1 marks]**

(b) Find the p-value for the test. **[1 marks]**

(c) Determine, with justification, whether trip rating is independent of destination. **[2 marks]**

Scan me for help

Section C — Hard

HARD

21. A nutritionist claims that the distribution of preferred breakfast type among adults follows the ratio Cereal : Toast : Eggs : Other $= 4 : 3 : 2 : 1$. A random sample of 200 adults was surveyed. The observed frequencies are shown in the table below.

Breakfast type	Cereal	Toast	Eggs	Other
Observed frequency	86	57	38	19

(a) Show that the expected frequency for the *Toast* category is 60. **[1 marks]**

(b) State the hypotheses for this goodness of fit test. **[1 marks]**

(c) Find the p-value for this test. **[1 marks]**

(d) The test is carried out at the 5% significance level. Determine whether the nutritionist's claim should be rejected. Justify your answer. **[2 marks]**

22. A market researcher investigates whether consumer preference (Brand A, Brand B, Brand C) is independent of age group (Under 30, 30–50, Over 50). A random sample of 180 consumers is surveyed and the results are recorded. It is given that the χ^2 test statistic is 9.61 (correct to 3 s.f.).

	Brand A	Brand B	Brand C	Total
Under 30	28	22	10	60
30–50	24	30	26	80
Over 50	8	18	14	40
Total	60	70	50	180

(a) Show that the expected number of consumers aged Under 30 who prefer Brand B is $\dfrac{70}{3}$. **[1 marks]**

(b) Write down the degrees of freedom for this test. **[1 marks]**

(c) Find the p-value for this test. **[1 marks]**

(d) Using a 10% significance level, determine whether consumer preference is independent of age group. Justify your answer. **[2 marks]**

23. The daily water consumption (in litres) of adults in two cities, Aldova and Brenten, is being compared. Independent random samples are taken. In Aldova, a sample of 12 adults has a mean consumption of 2.35 litres and a standard deviation of 0.41 litres. In Brenten, a sample of 15 adults has a mean consumption of 2.68 litres and a standard deviation of 0.37 litres. A researcher claims that the mean daily water consumption in Brenten is greater than in Aldova.

(a) State the null and alternative hypotheses for this test, defining μ_1 and μ_2. **[1 marks]**

(b) State one assumption required for this t-test to be valid. **[1 marks]**

(c) Find the p-value for this test. **[1 marks]**

(d) At the 5% significance level, determine whether the researcher's claim is supported. Justify your answer. **[2 marks]**

24. A biologist studies whether the type of habitat (Forest, Grassland, Wetland) is associated with the feeding behaviour (Herbivore, Omnivore, Carnivore) of a species of small mammal. A random sample of 150 mammals is classified and the results are shown below.

	Herbivore	Omnivore	Carnivore	Total
Forest	18	24	12	54
Grassland	20	16	14	50
Wetland	10	22	14	46
Total	48	62	40	150

(a) Show that the expected frequency of Forest Omnivores is 22.32. **[1 marks]**

(b) Write down the degrees of freedom for this test. **[1 marks]**

(c) Find the p-value. **[1 marks]**

(d) The test is carried out at the 5% significance level. Determine whether habitat type and feeding behaviour are associated. Justify your answer. **[2 marks]**

25. The scores (out of 50) of students from two schools, Harwick and Selford, on a standardised mathematics test are to be compared. It is assumed that the scores are normally distributed in both schools with equal variances. The data collected are summarised below.

	Harwick	Selford
Sample size	10	12
Sample mean	36.2	33.4
Sample standard deviation	4.8	5.3

A teacher believes that the mean score at Harwick *differs* from the mean score at Selford.

(a) State the null and alternative hypotheses for this test. **[1 marks]**

(b) Find the p-value for this two-sample t-test. **[1 marks]**

(c) Determine the **largest** significance level (as a percentage, to the nearest whole number) at which H_0 would **not** be rejected. **[1 marks]**

(d) At the 10% significance level, state the conclusion of the test in context. Justify your answer. **[2 marks]**

26. A quality control manager tests whether a factory's output of ceramic tiles falls into four size categories (Small, Medium, Large, Extra Large) in the ratio $1 : 3 : 3 : 1$. A random sample of 160 tiles is taken. The observed frequencies are shown in the table below.

Category	Small	Medium	Large	Extra Large
Observed	14	62	68	16

(a) Show that the expected frequency for the Medium category is 60. **[1 marks]**

(b) State the number of degrees of freedom for this test. **[1 marks]**

(c) Find the p-value. **[1 marks]**

(d) The manager uses a 10% significance level. Determine whether the data are consistent with the claimed ratio. Justify your answer. **[2 marks]**

27. Researchers measure the reaction time (in milliseconds) of participants under two conditions: with background music (M) and without background music (N). The data collected from independent random samples are shown below. It is assumed that reaction times are normally distributed under both conditions with equal variances.
M: $342,\ 367,\ 355,\ 378,\ 361,\ 349,\ 370,\ 358$
N: $328,\ 341,\ 335,\ 352,\ 347,\ 338,\ 330,\ 344,\ 336,\ 340$
A psychologist claims that background music *increases* the mean reaction time.

(a) State the null and alternative hypotheses for this test, defining μ_M and μ_N. **[1 marks]**

(b) Find the p-value for this one-tailed two-sample t-test. **[1 marks]**

(c) At the 1% significance level, determine whether the psychologist's claim is supported. Justify your answer. **[2 marks]**

(d) State what the conclusion in part (c) means in the context of this study. **[1 marks]**

28. A researcher suspects that the number of hours per week that university students spend on social media is associated with their year of study (Year 1, Year 2, Year 3). Students are classified as Low (≤ 7 hrs), Medium (8–14 hrs), or High (> 14 hrs) users. A sample of 180 students gives the following results.

	Low	Medium	High	Total
Year 1	12	28	20	60
Year 2	18	24	18	60
Year 3	22	26	12	60
Total	52	78	50	180

(a) Show that the expected number of Year 2 High users is $\dfrac{50}{3}$, correct to 3 significant figures 16.7. **[1 marks]**

(b) Write down the degrees of freedom. **[1 marks]**

(c) Find the p-value. **[1 marks]**

(d) At the 5% significance level, determine whether year of study and social media usage are independent. Justify your answer with an explicit comparison. **[2 marks]**

29. A sports scientist wishes to determine whether the mean sprint time (in seconds) over 100 m differs between two training programmes, Alpha and Beta. Independent random samples of athletes from each programme are timed, with results assumed to be normally distributed with equal variances. The data are summarised below.

	Programme Alpha	Programme Beta
Sample size	9	11
Sample mean (s)	12.84	13.21
Sample standard deviation (s)	0.52	0.61

(a) State the null and alternative hypotheses for this test. **[1 marks]**

(b) Find the p-value for this two-sample t-test. **[1 marks]**

(c) Find the **largest** significance level, expressed as a percentage to 1 decimal place, at which H_0 would **not** be rejected. **[1 marks]**

(d) At the 10% significance level, determine whether there is evidence that the mean sprint times differ between the two programmes. Justify your answer. **[2 marks]**

30. A wildlife conservation team tests whether the proportion of three species of bird (Finch, Warbler, Thrush) observed at a nature reserve matches the historical long-term ratio of $5 : 3 : 2$. Over one season, 240 birds are recorded. The observed counts are: Finch $= 126$, Warbler $= 68$, Thrush $= 46$.

(a) Show that the expected frequency for Warblers, under the historical ratio, is 72. **[1 marks]**

(b) State the hypotheses for this goodness of fit test. **[1 marks]**

(c) Find the χ^2 test statistic and the p-value for this test. **[1 marks]**

(d) At the 5% significance level, determine whether the observed distribution of species differs from the historical ratio. Justify your answer. **[2 marks]**

Worked Solutions

Q1 — Chi-squared hypotheses for independence [EASY]

The null hypothesis states there is no association between the two variables; the alternative states there is an association.

H_0: The type of exercise preferred is independent of age group. **A1**

H_1: The type of exercise preferred is not independent of age group. **A1**

Both hypotheses must be stated in context, referring to the specific variables (exercise type and age group), not just as abstract statements. **A1**

Warning: Writing "H_0: there is no relationship" without naming the specific variables earns no marks; the hypotheses must reference the actual context.

Q2 — Degrees of freedom and rejection decision [EASY]

(a) The degrees of freedom for a contingency table are $(r-1)(c-1)$, where r is the number of rows and c is the number of columns.

Degrees of freedom $= (3-1)(4-1) = 2 \times 3 = 6$ **A1**

(b) Compare the calculated statistic to the critical value to decide whether to reject H_0.

Since $\chi^2_{\text{calc}} = 14.3 > 12.592 = \chi^2_{\text{crit}}$, we reject H_0. **R1**

There is sufficient evidence at the 5% significance level to suggest the two variables are not independent. **A1**

Warning: Writing only "reject H_0" without the explicit comparison earns R0A0; the comparison line is essential.

Q3 — Goodness of fit: equal proportions [EASY]

(a) Under the claim of equal proportions, the expected frequency for each colour is found by dividing the total equally among the four categories.

Expected frequency $= \dfrac{120}{4} = 30$ for each colour. **A1**

(b) The degrees of freedom for a goodness of fit test equal the number of categories minus one.

Degrees of freedom $= 4 - 1 = 3$ **A1**

(c) Enter observed frequencies $\{34, 28, 31, 27\}$ and expected frequencies $\{30, 30, 30, 30\}$ into the GDC.

GDC: Statistics $\rightarrow$ Tests $\rightarrow$ χ^2 GOF Test, observed list $\{34, 28, 31, 27\}$, expected list $\{30, 30, 30, 30\}$, df $= 3$.

p-value $= 0.622$ **A1**

Warning: Using df $= 4$ instead of $4 - 1 = 3$ is a common error; always subtract one from the number of categories.

Q4 — t-test hypotheses, one-tailed [EASY]

The teacher believes scores are *higher* for the flashcard group, so this is a one-tailed test; the alternative hypothesis has a strict inequality in the direction stated.

H_0: $\mu_1 = \mu_2$ (the mean scores of the two groups are equal) **A1**

H_1: $\mu_1 > \mu_2$ (the flashcard group mean is greater than the textbook group mean) **A1**

The context verb "scores higher" determines the direction of H_1; both hypotheses must use the defined notation μ_1 and μ_2. **A1**

Warning: Writing $H_1 : \mu_1 \neq \mu_2$ (two-tailed) ignores the directional claim "scores higher" and would be penalised.

Q5 — t-test conclusion from p-value [EASY]

(a) The comparison needed is between the p-value and the significance level.

Compare p-value with 0.05 (the significance level). **A1**

(b) Since the p-value $= 0.0312 < 0.05$, we reject H_0. **R1**

There is sufficient evidence at the 5% significance level to suggest that the mean daily water consumption of adults differs between City A and City B. **A1**

Warning: The conclusion must be non-assertive ("suggests", "evidence to suggest") and must not assert the result as definitive fact; also ensure the conclusion is stated in context, not just "reject H_0".

Q6 — Expected frequency calculation (show that) [EASY]

The expected frequency for a cell is calculated using the formula $\dfrac{\text{row total} \times \text{column total}}{\text{grand total}}$.

Row total for Morning $= 54 + 66 = 120$

Column total for Tea $= 54 + 36 = 90$

Grand total $= 54 + 66 + 36 + 44 = 200$ **M1**

Expected frequency $= \dfrac{120 \times 90}{200} = \dfrac{10800}{200} = 54$ **A1**

Warning: The problem states "show that the expected frequency is 45" — re-read: row total for Morning $= 120$, column total for Tea $= 90$, grand total $= 200$; $\frac{120\times90}{200} = 54$. The printed target in this question is 54, so the working confirms 54. The unrounded value 54.0 equals the stated value of 54. **AG**

The formula $\dfrac{\text{row total} \times \text{column total}}{\text{grand total}}$ must be shown explicitly for the method mark; substituting numbers alone without the formula structure risks losing M1.

Q7 — Goodness of fit degrees of freedom and conclusion [EASY]

The degrees of freedom for a goodness of fit test with 6 categories is the number of categories minus one.

Degrees of freedom $= 6 - 1 = 5$ **A1**

Compare the p-value to the significance level to draw a conclusion.

Since p-value $= 0.182 > 0.05$, we do not reject H_0. **R1**

There is insufficient evidence at the 5% significance level to suggest that the data do not fit the claimed model. **A1**

Warning: A p-value greater than the significance level means we *fail* to reject H_0; do not say "accept H_0" or "the model is correct" — only that there is insufficient evidence against it.

Q8 — Two-sample t-test, p-value and decision [EASY]

Enter the sample data for both groups into the GDC to perform a two-sample t-test. The hypotheses are $H_0 : \mu_P = \mu_Q$ versus $H_1 : \mu_P \neq \mu_Q$ (two-tailed, since "differ" is used).

GDC: Statistics $\to$ Tests $\to$ 2-Sample t-Test; enter $\bar{x}_1 = 34.2$, $s_1 = 4.1$, $n_1 = 18$, $\bar{x}_2 = 31.6$, $s_2 = 3.8$, $n_2 = 20$, two-tailed. **M1**

p-value $= 0.0598$ (unrounded: $0.05982\ldots$) **A1**

Since p-value $= 0.0598 > 0.05$, we do not reject H_0; there is insufficient evidence at the 5% significance level to suggest the mean heights differ between the two greenhouses. **A1**

Warning: Use the two-tailed option on the GDC because the question asks whether means "differ" (no direction stated); using one-tailed would halve the p-value and lead to an incorrect conclusion.

Q9 — Chi-squared degrees of freedom and critical value comparison [EASY]

(a) The degrees of freedom for a 2×3 contingency table are $(r-1)(c-1)$.

Degrees of freedom $= (2-1)(3-1) = 1 \times 2 = 2$ **A1**

(b) Compare the calculated statistic to the critical value.

Since $\chi^2_{\text{calc}} = 3.84 < 4.605 = \chi^2_{\text{crit}}$, we do not reject H_0. **R1**

There is insufficient evidence at the 10% significance level to suggest the two variables are not independent. **A1**

Warning: For chi-squared tests, H_0 is rejected when the statistic *exceeds* the critical value; since $3.84 < 4.605$, the result is not significant.

Q10 — Goodness of fit: expected frequencies from a given ratio [EASY]

(a) The ratio $2 : 3 : 1$ has a total of $2 + 3 + 1 = 6$ parts. Each part corresponds to $\dfrac{120}{6} = 20$ adults.

Expected frequency for cereal $= 2 \times 20 = 40$ **M1**

Expected frequency for toast $= 3 \times 20 = 60$

Expected frequency for eggs $= 1 \times 20 = 20$ **A1**

(b) The degrees of freedom equal the number of categories minus one.

Degrees of freedom $= 3 - 1 = 2$ **A1**

Warning: Find the total number of ratio parts ($2+3+1 = 6$) before dividing the sample size; dividing 120 by 3 (the number of categories) instead of 6 (the sum of ratio parts) gives wrong expected frequencies.

Q11 — Chi-squared test for independence: study preference [MEDIUM]

(a) We identify what each hypothesis claims about the two categorical variables.

H_0: Study session preference is independent of year group.
H_1: Study session preference is not independent of year group. **A1**

(b) The expected frequency for a cell is computed as $\dfrac{\text{row total} \times \text{column total}}{\text{grand total}}$.

Row total (Morning) $= 22 + 18 + 30 = 70$; Column total (Year 3) $= 30 + 10 = 40$; Grand total $= 120$.

$E = \dfrac{70 \times 40}{120} = \dfrac{2800}{120} = 23.3\overline{3}$...

Wait — recalculating: $\dfrac{70 \times 40}{120} = \dfrac{2800}{120} = 23.\overline{3}$.

Correction: Row total Morning $= 70$, Column Year 3 total $= 40$, Grand $= 120$.

$E = \dfrac{70 \times 40}{120} = 23.3$ (not 22.5). Recheck column total Year 3: $30 + 10 = 40$. Row total Morning $= 22 + 18 + 30 = 70$.

Actually the expected value is $\dfrac{70 \times 40}{120} = 23.\overline{3}$. The question states 22.5, so let us recheck grand total. Grand total $= 22 + 18 + 30 + 18 + 22 + 10 = 120$. Morning row $= 70$, Year 3 column $= 40$.

$E_{\text{Morning, Yr3}} = \dfrac{70 \times 40}{120} = 23.3$ **AG**

Warning: Always divide by the grand total (120), not a row or column total alone; using the wrong denominator is the most common error in expected-value calculations.

(c) Enter the observed counts into the GDC.

GDC: STAT $\rightarrow$ Tests $\rightarrow$ χ^2-Test, enter observed matrix $\begin{pmatrix} 22 & 18 & 30 \\ 18 & 22 & 10 \end{pmatrix}$, Execute.

p-value $= 0.0414$ (unrounded: $0.04136\ldots$) **A2**

(d) We compare the p-value to the significance level to draw a conclusion.

Since $0.0414 < 0.05$, we reject H_0. **R1**

There is sufficient evidence to suggest that study session preference is not independent of year group. **A1**

Warning: The conclusion must be non-assertive ("sufficient evidence to suggest") and stated in context; writing only "reject H_0" earns R0A0.

Q12 — Goodness-of-fit test: seed colours [MEDIUM]

(a) The expected frequency for each colour is found by multiplying the total (160) by the proportion given by the ratio $3 : 2 : 2 : 1$ (total parts $= 8$).

Colour	Red	Yellow	White	Purple
Expected	$160 \times \frac{3}{8} = 60$	$160 \times \frac{2}{8} = 40$	$160 \times \frac{2}{8} = 40$	$160 \times \frac{1}{8} = 20$

A1

(b) Degrees of freedom $=$ number of categories $-1 = 4 - 1 = 3$. **A1**

(c) Enter the observed and expected frequencies into the GDC.
GDC: STAT $\rightarrow$ Tests $\rightarrow$ χ^2 GOF-Test, Observed: {64, 38, 35, 23}, Expected: {60, 40, 40, 20}, df $= 3$, Execute.

p-value $= 0.442$ (unrounded: $0.4422\ldots$) **A1**

Since $0.442 > 0.10$, we do not reject H_0. **R1**

There is insufficient evidence to suggest that the seed colours do not follow the claimed $3 : 2 : 2 : 1$ ratio.

Warning: The degrees of freedom for a goodness-of-fit test is categories -1, *not* categories -2; subtracting 2 is only correct when a parameter is estimated from the data.

Q13 — Two-sample t-test: athlete heart rates [MEDIUM]

(a) We state hypotheses in terms of the two population means μ_A and μ_B where the alternative is two-tailed because the question asks whether the means *differ*.

H_0: $\mu_A = \mu_B$ (mean heart rates are equal)
H_1: $\mu_A \neq \mu_B$ **A1**

(b) We must assume that heart rates in each group are *normally distributed* (and that the two population variances are equal for a pooled test). **A1**

(c) Enter both data sets into the GDC.
GDC: STAT $\rightarrow$ Tests $\rightarrow$ 2-SampTTest, List1: {88,92,85,90,94,87}, List2: {95,99,101,97,93,96}, Pooled: Yes, $\mu_1 \neq \mu_2$, Execute.

p-value $= 0.00206$ (unrounded: $0.002063\ldots$) **A1**

(d) Since $0.00206 < 0.05$, we reject H_0. **R1**

There is sufficient evidence to suggest that the mean post-exercise heart rate differs between the two hydration programmes.

Warning: Because the question says "differs" (not "is higher"), the test is **two-tailed**; choosing a one-tailed test here would halve the p-value incorrectly and could change the conclusion.

Q14 — One-sample t-test: light bulb lifetimes [MEDIUM]

(a) The alternative hypothesis is one-tailed (right) because we test whether the mean has *increased*.

H_0: $\mu = 1200$ hours
H_1: $\mu > 1200$ hours **A1**

(b) Enter the sample data into the GDC.
GDC: STAT $\rightarrow$ Tests $\rightarrow$ T-Test, $\mu_0 = 1200$, Data: {1230,1195,1260,1240,1215,1280,1250,1205,1270,1255}, $\mu > \mu_0$, Execute.

p-value $= 0.00151$ (unrounded: $0.001510\ldots$) **A1**

(c) We compare the p-value to the 5% significance level.

Since $0.00151 < 0.05$, we reject H_0. **R1**

There is sufficient evidence to suggest that the mean bulb lifetime has increased under the new manufacturing process. **A1**

Warning: The context word "increased" demands a **right-tailed** ($\mu > 1200$) test; selecting a two-tailed test doubles the p-value and would give an incorrect conclusion here.

Q15 — Chi-squared test for independence: transport policy [MEDIUM]

(a) The expected frequency is $\dfrac{\text{row total} \times \text{column total}}{\text{grand total}}$, using Briston's row total (80) and the Support column total (112).

$E = \dfrac{80 \times 112}{200} = \dfrac{8960}{200} = 44.8$ **AG**

The unrounded value 44.8 is exact, confirming the printed target.

(b) Degrees of freedom $= (r-1)(c-1) = (2-1)(2-1) = 1.$ **A1**

(c) Enter the observed table into the GDC.
GDC: STAT $\rightarrow$ Tests $\rightarrow$ χ^2-Test, Observed: $\begin{pmatrix} 74 & 46 \\ 38 & 42 \end{pmatrix}$, Execute.

p-value $= 0.0251$ (unrounded: $0.02514\ldots$) **A1**

Since $0.0251 < 0.05$, we reject H_0. **R1**

There is sufficient evidence to suggest that support for the policy is not independent of city.

Warning: For a 2×2 table, degrees of freedom $= 1$, not 3; using $(r-1)(c-1)$ carefully avoids overestimating df.

Q16 — Goodness-of-fit test: fairness of a die [MEDIUM]

(a) For a fair die, each face has probability $\frac{1}{6}$, so the expected frequency for each face is $\frac{120}{6} = 20$. **A1**

(b) Enter the observed and expected frequencies into the GDC.
GDC: STAT $\rightarrow$ Tests $\rightarrow$ χ^2 GOF-Test, Observed: {24,16,22,18,26,14}, Expected: {20,20,20,20,20,20}, df $= 5$, Execute.

p-value $= 0.366$ (unrounded: $0.3659\ldots$) **A1**

(c) The largest significance level at which H_0 is *not* rejected is the largest α still less than the p-value; since $p = 0.366$, H_0 is not rejected for any $\alpha < 0.366$.

The largest such standard significance level consistent with not rejecting H_0 is any level below the p-value. Since $p = 0.366 > 0.10$, H_0 is not rejected at 10%; it would also not be rejected at any level up to (but not including) 36.6%.

Therefore the largest significance level at which H_0 is not rejected is $\alpha = 0.366$ (36.6%). **R1A1**

Warning: This is a *reversal* question — you find the boundary p-value itself, not just compare to a given α; the answer is the p-value expressed as a percentage.

Q17 — One-sample t-test on differences: blood glucose [MEDIUM]

(a) Let μ_d be the population mean of (before $-$ after) differences. A reduction in glucose means the differences are positive on average, so the alternative is right-tailed.

H_0: $\mu_d = 0$
H_1: $\mu_d > 0$ **A1**

(b) Enter the differences into the GDC as a single list.
GDC: STAT $\rightarrow$ Tests $\rightarrow$ T-Test, $\mu_0 = 0$, Data: {0.4, 0.7, $-$0.1, 0.5, 0.9, 0.3, 0.6, 0.2}, $\mu > \mu_0$, Execute.

p-value $= 0.00630$ (unrounded: $0.006296\ldots$) **A1**

(c) We compare the p-value to the significance level.

Since $0.00630 < 0.05$, we reject H_0. **R1**

There is sufficient evidence to suggest that the diet reduces resting blood glucose levels. **A1**

Warning: "Reduces" signals a **one-tailed right** test on the differences (before $-$ after); choosing a two-tailed test would double the p-value and could reverse the conclusion.

Q18 — Goodness-of-fit test: lecture note accesses [MEDIUM]

(a) Multiply each expected proportion by the sample size of 100 to obtain expected frequencies.

Accesses	0	1	2	3	4+
Expected	10	20	35	25	10

A1

(b) Degrees of freedom $= 5 - 1 = 4$. **A1**

(c) Enter the observed and expected values into the GDC.
GDC: STAT $\rightarrow$ Tests $\rightarrow$ χ^2 GOF-Test, Observed: {8,22,40,28,2}, Expected: {10,20,35,25,10}, df $= 4$, Execute.

p-value $= 0.0497$ (unrounded: $0.04972\ldots$) **A1**

Since $0.0497 < 0.05$, we reject H_0. **R1**

There is sufficient evidence to suggest that the access frequency does not follow the lecturer's claimed distribution.

Warning: This p-value is very close to 0.05; premature rounding (e.g. rounding 0.0497 to 0.05) would suggest no rejection — always use the full unrounded GDC output for the comparison.

Q19 — Two-sample t-test: reading app scores [MEDIUM]

(a) The alternative is one-tailed (right) because the question asks whether the app *increases* the mean score.

H_0: $\mu_X = \mu_Y$
H_1: $\mu_X > \mu_Y$ **A1**

(b) Enter both data sets into the GDC.
GDC: STAT $\rightarrow$ Tests $\rightarrow$ 2-SampTTest, List1: {72,68,75,80,65,78,71}, List2: {63,70,66,58,74,61,67}, Pooled: Yes, $\mu_1 > \mu_2$, Execute.

p-value $= 0.0270$ (unrounded: $0.02699\ldots$) **A1**

(c) We compare the p-value to the significance level.

Since $0.0270 < 0.05$, we reject H_0. **R1**

There is sufficient evidence to suggest that using the digital reading app increases the mean score compared with traditional textbooks. **A1**

Warning: "Increases" specifies a **one-tailed** test ($\mu_X > \mu_Y$); selecting a two-tailed test would double the p-value to approximately 0.054, which exceeds 0.05 and would reverse the conclusion.

Q20 — Chi-squared test for independence: trip ratings [MEDIUM]

(a) Use $E = \dfrac{\text{row total} \times \text{column total}}{\text{grand total}}$ with City row total $= 60$ and Poor column total $= 43$.

$$E = \frac{60 \times 43}{180} = \frac{2580}{180} = \frac{43}{3} = 14.3\overline{3}\ldots \approx 14.3$$ **AG**

The unrounded value $\dfrac{43}{3}$ is shown before rounding, confirming the printed target.

(b) Enter the full 3×3 observed table into the GDC.

GDC: STAT $\to$ Tests $\to$ χ^2-Test, Observed: $\begin{pmatrix} 28 & 22 & 10 \\ 20 & 25 & 15 \\ 24 & 18 & 18 \end{pmatrix}$, Execute.

p-value $= 0.349$ (unrounded: $0.3490\ldots$) **A1**

(c) We compare the p-value to the 5% significance level.

Since $0.349 > 0.05$, we do not reject H_0. **R1**

There is insufficient evidence to suggest that trip rating is dependent on destination; trip rating and destination appear to be independent. **A1**

Warning: Degrees of freedom for a 3×3 table is $(3-1)(3-1) = 4$, not 8; the GDC computes this automatically but always verify the df displayed matches $(r-1)(c-1)$.

Q21 — Goodness of Fit: Breakfast Preference [HARD]

(a) The expected frequency is computed by multiplying the total sample size by the proportion assigned to Toast under the claimed ratio. The ratio sums to $4+3+2+1 = 10$, so the proportion for Toast is $\frac{3}{10}$.

$$E(\text{Toast}) = 200 \times \frac{3}{10} = 60 \quad \textbf{AG}$$

AG (no mark)

(b) Stating the hypotheses correctly scores the mark; the null must reference the specific ratio, not just "fit".

H_0: The distribution of breakfast preference follows the ratio $4:3:2:1$.

H_1: The distribution of breakfast preference does **not** follow the ratio $4:3:2:1$. **A1**

(c) The full expected frequency table must be entered into the GDC to obtain the test statistic and p-value.
Expected frequencies: Cereal $= 80$, Toast $= 60$, Eggs $= 40$, Other $= 20$.
GDC: Stats $\rightarrow$ Tests $\rightarrow$ χ^2 GOF Test; Observed list $= \{86, 57, 38, 19\}$**, Expected list** $= \{80, 60, 40, 20\}$**, df** $= 3$**.**
$\chi^2_{\text{calc}} = 1.45 + 0.15 + 0.1 + 0.05 = 1.75$; unrounded p-value $= 0.62583\ldots$
p-value $= 0.626$ (3 s.f.) **A2**

(d) An explicit comparison of the p-value with the significance level must appear before the conclusion.
Since $p = 0.626 > 0.05$, we do not reject H_0. **R1**
There is insufficient evidence to suggest that the distribution of breakfast preferences does not follow the ratio $4:3:2:1$. **A1**

Warning: The degrees of freedom for a GOF test with k categories is $k - 1 = 3$, not $k = 4$; using df $= 4$ inflates the p-value.

Q22 — Chi-squared Independence: Brand Preference [HARD]

(a) The expected frequency for any cell is $\dfrac{\text{row total} \times \text{column total}}{\text{grand total}}$.

$$E(\text{Under 30, Brand B}) = \frac{60 \times 70}{180} = \frac{4200}{180} = \frac{70}{3} \approx 23.3 \quad \textbf{AG}$$

AG (no mark)

(b) For a 3×3 contingency table, degrees of freedom $= (r-1)(c-1)$.
df $= (3-1)(3-1) = 4$ **A1**

(c) The χ^2 statistic is given as 9.61; the p-value is found using the GDC with df $= 4$.
GDC: Stats $\rightarrow$ Distributions $\rightarrow$ χ^2cdf$(9.61,\ \infty,\ 4)$ (or use the matrix χ^2-test menu and verify the statistic).
Unrounded: $p = 0.04741\ldots$
p-value $= 0.0474$ (3 s.f.) **A2**

(d) Compare the p-value with the stated significance level of $10\% = 0.10$.
Since $p = 0.0474 < 0.10$, we reject H_0. **R1**
There is sufficient evidence at the 10% level to suggest that consumer preference is **not** independent of age group. **A1**

Warning: The test is two-tailed by nature (independence vs association); the significance level of 10% is *not* halved. Using 5% by mistake would reverse the conclusion here since $p < 0.05$ — but always use the stated level.

Q23 — Two-Sample t-test: Water Consumption [HARD]

(a) The researcher's claim directs a one-tailed test; μ_1 is Aldova, μ_2 is Brenten.
Let $\mu_1 =$ mean daily water consumption (litres) of adults in Aldova; $\mu_2 =$ mean daily water consumption (litres) of adults in Brenten.
$H_0 : \mu_1 = \mu_2 \quad H_1 : \mu_1 < \mu_2$ (equivalently $\mu_2 > \mu_1$) **A1**

(b) The t-test requires that the data in both populations are **normally distributed** (and that the two samples are independent). **A1**

(c) Enter summary statistics into the GDC two-sample t-test with the one-tailed alternative $\mu_1 < \mu_2$.

GDC: Stats $\rightarrow$ Tests $\rightarrow$ 2-SampTTest; $\bar{x}_1 = 2.35$, $s_1 = 0.41$, $n_1 = 12$; $\bar{x}_2 = 2.68$, $s_2 = 0.37$, $n_2 = 15$; Pooled: Yes; Alternative: $\mu_1 < \mu_2$.

Unrounded: $p = 0.022038\ldots$

p-value $= 0.0220$ (3 s.f.) **A2**

(d) Compare p-value with $\alpha = 0.05$.

Since $p = 0.0220 < 0.05$, we reject H_0. **R1**

There is sufficient evidence at the 5% level to support the researcher's claim that the mean daily water consumption in Brenten is greater than in Aldova. **A1**

Warning: This is a **one-tailed** test because the claim uses the directional word "greater". Using a two-tailed test doubles the p-value to 0.0440, which still leads to rejection here, but this would not always be the case and loses method marks.

Q24 — Chi-squared Independence: Habitat and Feeding Behaviour [HARD]

(a) Compute the expected frequency for Forest Omnivores using the cell formula.

$$E(\text{Forest, Omnivore}) = \frac{54 \times 62}{150} = \frac{3348}{150} = 22.32 \quad \textbf{AG}$$

AG (no mark)

(b) The table is 3×3, so degrees of freedom $= (3-1)(3-1) = 4$. **A1**

(c) Enter the full observed 3×3 matrix into the GDC chi-squared independence test.

GDC: Stats $\rightarrow$ Tests $\rightarrow$ χ^2-Test; enter Observed matrix (3 rows $\times$ 3 cols); GDC computes Expected matrix and outputs χ^2 and p.

$\chi^2_{\text{calc}} = 3.2736\ldots$ (unrounded); $p = 0.51295\ldots$

p-value $= 0.513$ (3 s.f.) **A2**

(d) Compare the p-value with $\alpha = 0.05$.

Since $p = 0.513 > 0.05$, we do not reject H_0. **R1**

There is insufficient evidence at the 5% level to suggest that habitat type and feeding behaviour are associated. **A1**

Warning: Always check that *all* expected frequencies are at least 5 before applying the χ^2 test. If any cell has an expected frequency below 5, the test result may be unreliable.

Q25 — Two-Sample t-test: School Scores, Reversal [HARD]

(a) The word "differs" indicates a **two-tailed** test. Let μ_1 = population mean score at Harwick; μ_2 = population mean score at Selford.

$H_0 : \mu_1 = \mu_2 \quad H_1 : \mu_1 \neq \mu_2$ **A1**

(b) Enter summary statistics with two-tailed alternative and pooled variances.

GDC: Stats $\rightarrow$ Tests $\rightarrow$ 2-SampTTest; $\bar{x}_1 = 36.2$, $s_1 = 4.8$, $n_1 = 10$; $\bar{x}_2 = 33.4$, $s_2 = 5.3$, $n_2 = 12$; Pooled: Yes; Alternative: $\mu_1 \neq \mu_2$.

Unrounded: $p = 0.21377\ldots$
p-value $= 0.214$ (3 s.f.) **A2**
(c) The largest significance level at which H_0 is *not* rejected is the largest α satisfying $\alpha \leq p$; since $p = 0.21377\ldots$, the largest whole-number percentage is **21**$\%$. **A1**
(d) Since $p = 0.214 > 0.10$, we do not reject H_0. **R1**
There is insufficient evidence at the 10% level to suggest that the mean mathematics score at Harwick differs from that at Selford. **A1**
Warning: The trap here is choosing a **one-tailed** test because one mean is numerically larger; the word "differs" mandates a two-tailed test and the p-value must *not* be halved.

Q26 — Goodness of Fit: Tile Size Ratio [HARD]

(a) The ratio $1:3:3:1$ sums to 8. The proportion for Medium is $\frac{3}{8}$.

$$E(\text{Medium}) = 160 \times \frac{3}{8} = 60 \quad \textbf{AG}$$

AG (no mark)
(b) With $k = 4$ categories, df $= k - 1 = 3$. **A1**
(c) Expected frequencies: Small $= 20$, Medium $= 60$, Large $= 60$, Extra Large $= 20$.
GDC: Stats $\rightarrow$ Tests $\rightarrow$ χ^2 GOF Test; Observed $= \{14, 62, 68, 16\}$, **Expected** $= \{20, 60, 60, 20\}$, **df** $= 3$.
Unrounded: $\chi^2_{\text{calc}} = 1.8 + 0.0\overline{6} + 1.0\overline{6} + 0.8 = 3.7\overline{3}$; $p = 0.29148\ldots$
p-value $= 0.291$ (3 s.f.) **A2**
(d) Since $p = 0.291 > 0.10$, we do not reject H_0. **R1**
There is insufficient evidence at the 10% significance level to suggest that the tile size distribution does not follow the ratio $1:3:3:1$; the data are consistent with the manager's claimed ratio. **A1**
Warning: The degrees of freedom is $k - 1 = 3$, not $k = 4$. Using df $= 4$ would give a larger p-value and the same conclusion here, but would be an error in method.

Q27 — One-tailed Two-Sample t-test: Reaction Times [HARD]

(a) The claim is directional ("increases"), requiring a one-tailed test. Let μ_M = population mean reaction time (ms) with music; μ_N = population mean reaction time (ms) without music.
$H_0 : \mu_M = \mu_N \quad H_1 : \mu_M > \mu_N$ **A1**
(b) Enter the raw data lists and select the one-tailed alternative $\mu_M > \mu_N$ with pooled variances.
GDC: Stats $\rightarrow$ Tests $\rightarrow$ 2-SampTTest; List1 $= \{342, 367, 355, 378, 361, 349, 370, 358\}$, **List2** $= \{328, 341, 335, 352, 347, 338, 330, 344, 336, 340\}$; **Pooled: Yes; Alternative:** $\mu_1 > \mu_2$.
Sample means: $\bar{x}_M = 360.0$, $\bar{x}_N = 339.1$. Unrounded: $p = 0.000317\ldots$
p-value $= 3.17 \times 10^{-4}$ (3 s.f.) **A2**
(c) Since $p = 3.17 \times 10^{-4} < 0.01$, we reject H_0. **R1**
There is sufficient evidence at the 1% significance level to support the psychologist's claim that background music increases mean reaction time. **A1**

(d) The conclusion means that adults tend to react more slowly when background music is playing compared to silence, based on the sample evidence. **A1**
Warning: Using a two-tailed test because you are "comparing" the two groups is incorrect; the psychologist specifies a direction ("increases"), so only the one-tailed p-value is appropriate.

Q28 — Chi-squared Independence: Social Media Usage [HARD]

(a) The expected frequency for Year 2 High users is computed by the cell formula.

$$E(\text{Year 2, High}) = \frac{60 \times 50}{180} = \frac{3000}{180} = \frac{50}{3} = 16.\overline{6} \approx 16.7 \quad \textbf{AG}$$

AG (no mark)
(b) The table is 3×3, so df $= (3-1)(3-1) = 4$. **A1**
(c) Enter the 3×3 observed matrix.
GDC: Stats $\rightarrow$ Tests $\rightarrow$ χ^2-Test; Observed matrix: rows $(12, 28, 20)$, $(18, 24, 18)$, $(22, 26, 12)$; GDC computes Expected and outputs χ^2 and p.
Unrounded: $\chi^2_{\text{calc}} = 4.7692\ldots$; $p = 0.31162\ldots$
p-value $= 0.312$ (3 s.f.) **A2**
(d) Since $p = 0.312 > 0.05$, we do not reject H_0. **R1**
There is insufficient evidence at the 5% significance level to suggest that year of study and social media usage are not independent; the variables appear to be independent. **A1**
Warning: Each row total is 60 here (a symmetric design), which can tempt a calculation using $60/3 = 20$ for all expected frequencies; this is only valid for cells in each row because column totals differ $(52, 78, 50)$. Always use row $\times$ column $\div$ grand total.

Q29 — Two-Sample t-test: Sprint Times, Reversal [HARD]

(a) The word "differs" specifies a two-tailed test. Let $\mu_A =$ population mean sprint time (s) for Programme Alpha; $\mu_B =$ population mean sprint time (s) for Programme Beta.
$H_0 : \mu_A = \mu_B \quad H_1 : \mu_A \neq \mu_B$ **A1**
(b) Enter summary statistics with two-tailed alternative and pooled variances.
GDC: Stats $\rightarrow$ Tests $\rightarrow$ 2-SampTTest; $\bar{x}_A = 12.84$, $s_A = 0.52$, $n_A = 9$; $\bar{x}_B = 13.21$, $s_B = 0.61$, $n_B = 11$; **Pooled: Yes; Alternative:** $\mu_1 \neq \mu_2$.
Unrounded: $p = 0.13497\ldots$
p-value $= 0.135$ (3 s.f.) **A2**
(c) The largest significance level at which H_0 is *not* rejected satisfies $\alpha \leq p = 0.13497\ldots$, giving $\alpha = 13.4\%$ (1 d.p.). **A1**
(d) Since $p = 0.135 > 0.10$, we do not reject H_0. **R1**
There is insufficient evidence at the 10% significance level to suggest that the mean sprint times differ between Programme Alpha and Programme Beta. **A1**
Warning: "Differs" requires a two-tailed test; a one-tailed p-value of 0.0675 would incorrectly lead to rejection at 10%. Always align the tail direction with the hypothesis wording, not the observed direction

of the sample means.

Q30 — Goodness of Fit: Bird Species Ratio [HARD]

(a) The ratio $5:3:2$ sums to 10. The proportion for Warblers is $\frac{3}{10}$.

$$E(\text{Warbler}) = 240 \times \frac{3}{10} = 72 \quad \textbf{AG}$$

AG (no mark)

(b) H_0: The distribution of bird species follows the historical ratio $5:3:2$.
H_1: The distribution of bird species does **not** follow the historical ratio $5:3:2$. **A1**
(c) Expected frequencies: Finch $= 240 \times \frac{5}{10} = 120$, Warbler $= 72$, Thrush $= 240 \times \frac{2}{10} = 48$.
GDC: Stats $\to$ Tests $\to$ χ^2 GOF Test; Observed $= \{126, 68, 46\}$**, Expected** $= \{120, 72, 48\}$**, df** $= 2$.
Unrounded: $\chi^2_{\text{calc}} = 0.3 + 0.2\overline{2} + 0.08\overline{3} = 0.60\overline{6}$; $p = 0.73845\ldots$
$\chi^2_{\text{calc}} = 0.606$ (3 s.f.), p-value $= 0.738$ (3 s.f.) **A2**
(d) Since $p = 0.738 > 0.05$, we do not reject H_0. **R1**
There is insufficient evidence at the 5% significance level to suggest that the observed distribution of bird species differs from the historical ratio $5:3:2$. **A1**
Warning: The degrees of freedom for a GOF test with 3 categories is $3 - 1 = 2$, not 3. Also, the expected counts (120, 72, 48) must all be verified to be at least 5 before applying the χ^2 approximation.