IB DP MATHS AI · TOPIC 4 STATISTICS AND PROBABILITY · SL 4.11 · ~6 MIN READ

Chi-squared and t-tests

Hypothesis testing for IB Maths AI: null and alternative hypotheses, significance levels and p-values, the χ^2 test for independence with contingency tables and degrees of freedom, the χ^2 goodness of fit test, and the two-sample t-test with one- and two-tailed alternatives, all with GDC steps and conclusions in context.

Null & alternative hypotheses | Significance level | p-values | Critical values | Contingency tables | Expected frequencies | Degrees of freedom | χ^2 test for independence

RS by Riyaz Sir
mentorriyaz.in

No tag = **SL and HL** HL ONLY HL students only FORMULA BOOKLET given in the exam

M1 A1 how marks are awarded highlight = comes up again and again Test yourself at the end

IN 30 SECONDS

- A **hypothesis test** starts from H_0 (no effect, no difference, independent) and asks whether the data is unlikely enough to reject it.
- **Decision rule:** if the **p-value < significance level** (usually 5%), reject H_0. Equivalently, reject if the test statistic is beyond the critical value.
- **χ^2 test for independence:** compare observed frequencies in a contingency table with the expected frequencies $E = \frac{\text{row total} \times \text{column total}}{\text{grand total}}$. Degrees of freedom = (rows − 1)(columns − 1).
- **χ^2 goodness of fit:** compare observed frequencies with those a model (uniform, binomial, normal) predicts. Degrees of freedom = number of categories − 1 (for the AI course, where no parameters are estimated).
- **t-test:** compares the means of two independent samples from normal populations. Use a one-tailed test when H_1 says "greater" or "less", two-tailed for "different".

BY THE END YOU CAN

- Write null and alternative hypotheses, and understand significance levels and p-values.
- Find expected frequencies and degrees of freedom for a contingency table.
- Carry out and interpret a χ^2 test for independence.
- Carry out and interpret a χ^2 goodness of fit test.
- Carry out and interpret a t-test for the difference of two means, one- or two-tailed.
- Write conclusions in context.

Riyaz Sir

1 The language of hypothesis tests

Term	Meaning
H_0	null hypothesis: no association, no difference
H_1	alternative hypothesis: what you suspect
significance level	the probability of rejecting H_0 when it is true (5% unless told)
p-value	the probability of a result at least this extreme if H_0 is true
critical value	the boundary of the rejection region for the test statistic

COMMON MISTAKE

Writing "accept H_0". The data never proves H_0; you **do not reject** it, because there is not enough evidence against it.

2 χ^2 test for independence

$$\chi^2_{calc} = \Sigma \frac{(f_o - f_e)^2}{f_e}$$ FORMULA BOOKLET

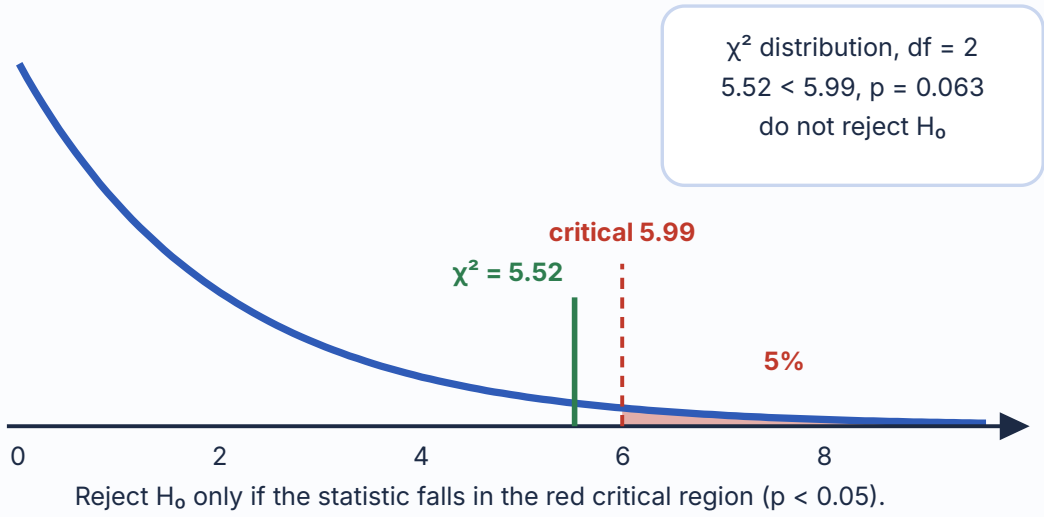
f_o observed, f_e expected frequency

	Maths	Science	Arts	Total
Year 11	30	25	15	70
Year 12	20	35	25	80
Total	50	60	40	150

WORKED EXAMPLE

Test at the 5% level whether favourite subject is independent of year group.

1. H_0: favourite subject and year group are independent; H_1: they are not independent **A1**
2. expected Year 11 Maths = $\frac{70 \times 50}{150}$ = 23.3, ...; df = (2 − 1)(3 − 1) = 2 **M1** **A1**
3. GDC χ^2 test: **χ^2 = 5.52, p = 0.0631** **A1**
4. **0.0631 > 0.05, so do not reject H_0: there is not enough evidence at 5% that favourite subject depends on year group** **R1**



Reject H_0 only if the statistic falls in the red critical region ($p < 0.05$).

The χ^2 distribution with 2 degrees of freedom. 5.52 falls just short of the 5% critical region.

EXAM TIP

Expected frequencies must be at least 5. If some are smaller, combine rows or columns (and recalculate the degrees of freedom).

GDC (CALCULATOR) METHOD

Put the observed table in matrix A; the χ^2 test puts the expected values in matrix B. Write down at least one expected value, the degrees of freedom, χ^2 and p.

3 χ^2 goodness of fit

WORKED EXAMPLE

A die is rolled 60 times: 1 (8), 2 (12), 3 (9), 4 (7), 5 (14), 6 (10). Test at 5% whether it is fair.

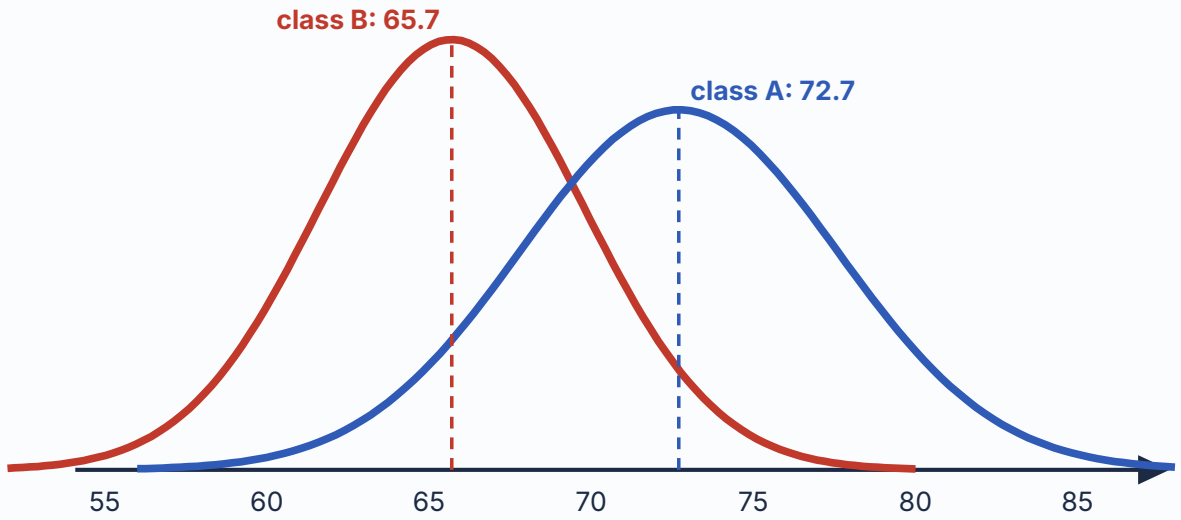
1. H_0: the die is fair (each face has probability $\frac{1}{6}$); H_1: it is not fair **A1**
2. expected 10 each, df = 6 − 1 = 5 **A1**
3. $\chi^2 = \frac{4 + 4 + 1 + 9 + 16 + 0}{10} = \mathbf{3.40}$, p = **0.639** **M1** **A1**
4. **p > 0.05 (and 3.40 < critical value 11.07): do not reject H_0; there is no evidence the die is biased** **R1**

- **Other models:** for "the data follows B(4, 0.5)" or a normal model, find the expected frequencies from the model × total, then test in the same way.

COMMON MISTAKE

Using percentages instead of frequencies. χ^2 uses counts: convert probabilities to expected **counts** by multiplying by the total.

4 The t-test for two means



t = 2.88, p = 0.0070 (one-tailed): the difference in means is significant at 5%.

Two sample means 7 marks apart. Is that difference real or just sampling variation?

WORKED EXAMPLE

Class A scored 72, 68, 75, 80, 66, 71, 77 and class B scored 65, 70, 62, 68, 64, 71, 60. Test at the 5% level whether class A's mean is greater. Assume normal populations with equal variances.

1. H_0: $\mu_A = \mu_B$; H_1: $\mu_A > \mu_B$ (one-tailed) **A1**
2. 2-sample t-test, pooled: $\bar{x}_A = 72.7$, $\bar{x}_B = 65.7$, **$t = 2.88$, $p = 0.00697$** **M1** **A1**
3. **$p < 0.05$, so reject H_0: there is evidence that class A's mean score is higher** **R1**

- **Two-tailed:** for H_1: $\mu_A \neq \mu_B$, the p-value doubles (here 0.0139), and you compare with 5% in the same way.
- **Assumptions:** both samples random and independent, both populations normally distributed. Mention them if asked to comment on the validity of the test.

LINKING QUESTION

The χ^2 test checks the models of Probability, binomial and normal distributions against data, and the t-test uses the ideas of sampling distributions from the central limit theorem (Estimation, Poisson, hypothesis tests and Markov chains, HL). Correlation in Statistics, correlation and regression tells you about a relationship; these tests tell you whether it is significant.

IDEA FOR YOUR IA / EXPLORATION

Survey students about two categorical variables (year group and preferred study method, gender and favourite sport) and run a χ^2 test for independence, discussing expected frequencies, combining categories and the limits of your sample. Or compare reaction times of two groups with a t-test.

Test yourself — cover the answers first

QUICK CHECK

1. What does H_0 say in a test for independence?
2. Find df for a 3 × 4 contingency table.
3. p = 0.03 at the 5% level. Decision?
4. Expected frequency: row total 40, column total 30, grand total 120.

5. Which H_1 gives a one-tailed t-test: "different" or "greater"?

EXAM-STYLE

6. In a goodness of fit test with 4 categories, $\chi^2_{calc} = 9.2$ and the 5% critical value for 3 df is 7.815. State and justify your conclusion.
7. A contingency table has an expected frequency of 3.2 in one cell. What should you do, and why?

Show answers

1. the two variables are independent
2. 6
3. reject H_o
4. 10
5. greater
6. $9.2 > 7.815$, so χ^2 is in the critical region
   reject H_o: there is evidence at the 5% level that the data does not fit the model
7. expected frequencies should be at least 5 for the χ^2 test to be valid
   combine that row or column with an adjacent one, then recalculate the degrees of freedom before testing

Stuck on a question?

Send it to me on WhatsApp. I read every message.

Share with a friend

Ask Riyaz Sir

Want regular help? Send this to your parent →