PHOTON ACADEMY · IB MATH AI HL CHEATSHEET

Hypothesis Testing

SL 4.11 · AHL 4.18–4.19

ibtuition.sg · free

FORMULAS YOU MUST KNOW

H_0 — "no effect": the variables are independent, the data fit the model, or $\mu = \mu_0$.

Null hypothesis:

H_1 — an effect exists: the variables are associated, the data do not fit, or the mean has changed.

Alternative:

α — the risk you accept of rejecting a true H_0; usually $\alpha = 0.05$ (5%), written as a decimal.

Significance level:

reject H_0 when $p < \alpha$; otherwise there is insufficient evidence to reject it.

Decision rule:

$E = n \times p$ — the sample size times the proportion H_0 predicts for that category.

From a proportion:

turn the % into a fraction over 100 first: $E = n \times \dfrac{\%}{100}$.

From a percentage:

$E = \dfrac{(\text{row total}) \times (\text{column total})}{\text{grand total}}$ — for a test of independence.

Contingency cell:

$\chi^2_{\text{calc}} = \sum \dfrac{(O-E)^2}{E}$ — summed over every cell or category.

Test statistic:

$\text{df} = k - 1$ for k categories.

df (goodness of fit):

$\text{df} = (r-1)(c-1)$ for an $r \times c$ table.

df (independence):

each $E = \dfrac{n}{k}$ — n data split evenly across k categories.

Equally likely:

$E_i = n \times p_i$ — scale each model proportion by the sample size.

Stated proportions:

KEY METHODS

- A test never proves H_0 true. You either reject it, or you fail to reject it (insufficient evidence). A 1% test is stricter than a 5% test — harder to reject H_0 — so always use the α the question states.
- Expected frequencies need not be whole numbers; keep them exact. If any expected frequency comes out below 5, combine adjacent categories before testing (and recount the number of categories for the degrees of freedom).
- Keep O and E paired: square each difference, divide by its own E, then add. Never divide the grand total by E at the end.
- Enter the observed matrix into your GDC's χ^2-test and it returns χ^2_{calc}, the p-value and the degrees of

EXAM TRAPS TO AVOID

- The claim you are testing for — a difference, an effect, an association — always goes in H_1. H_0 is the "nothing interesting" statement and always carries the equality.
- Build each expected cell from the totals around the edge of the table — the row total and column total — never from the observed value sitting in the cell.
- Degrees of freedom depend only on the table's shape, not the sample size — a 2×3 table always has $\text{df} = (2-1)(3-1) = 2$, whether you tested 40 people or 4000.
- A large χ^2 (small p-value) means the data do not fit — you reject H_0. Do not read a big statistic as

freedom in one step. Do the sum by hand only for a "show that" part.

- A quick self-check: each row and each column of your expected table must add up to the same totals as the observed table. If they do not, an expected value is wrong.
- The categories must together cover every outcome and their expected frequencies must total n. If a class has an expected count below 5, merge it with a neighbour before you test.
- Work left to right: find each $P(X = k)$ on the GDC (keeping the $e^{-\lambda}$ factor), turn it into a frequency with $E_k = nP(X = k)$, then form $\frac{(O-E)^2}{E}$ for each class.

evidence of a good fit; it is the opposite.

- The χ^2 test is one-tailed on the upper side only — you reject for a large statistic. There is no "too small to be true" rejection here.
- Whether the test is one- or two-tailed is fixed by H_1. A two-tailed test ($\neq$) splits α across both tails, so make sure the GDC's tail setting matches your H_1 — the wrong setting doubles or halves the p-value.
- Divide by $\sqrt{n}$, not n. The standard error shrinks only with the square root of the sample size, so quadrupling n merely halves it.