

O · Statistical hypothesis testing · supported focus

AQA · A-Level · Mathematics

Scope and prerequisites

AQA A-level Mathematics 7357, section O (Statistical hypothesis testing), version 1.3 (first assessment 2018). Original teaching notes; objective-level mapping lives in [content/teaching/aqa_a_level_mathematics/objective-review.yaml](content/teaching/aqa_a_level_mathematics/objective-review.yaml).

Prerequisites: the binomial and normal models from section N. Each lesson develops its own method before independent transfer.

These are original teaching examples, not official AQA material.

Statistical hypothesis testing

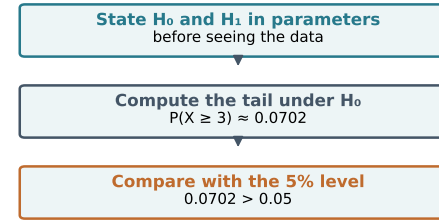
■ What would explain this observation?

- A factory claims a 10% defect rate, and a sample of 10 has 3 defects. A student says “3 defects proves the rate is above 10%.” What is the right way to test that claim?
- Start with a prediction. State the default claim, the alternative, and what probability the evidence needs to be judged against.

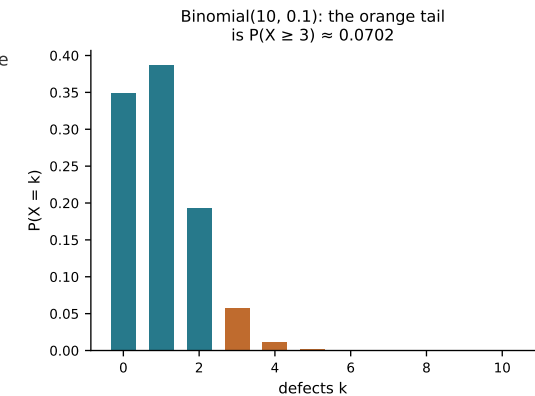
■ Build the model

- State the **null hypothesis** H_0 and the **alternative hypothesis** H_1 in population parameters BEFORE inspecting the outcome. Compute the probability, under H_0 , of a result at least as extreme as the observed one – the tail probability.
- Compare it with the chosen **significance level** (such as 5%). If the tail meets the level, reject H_0 ; otherwise say the evidence is INSUFFICIENT, never that H_0 is proven. Statistical significance says nothing about the practical SIZE of an effect.
- null hypothesis: The default claim about a population parameter, tested against the evidence; alternative hypothesis: The claim accepted if the evidence against H_0 is strong enough; significance level: The chosen probability threshold for rejecting the null hypothesis.

Test the claim by the probability of evidence THIS extreme



insufficient evidence $\neq$ proof that H_0 is true



■ Choose evidence that can test it

- Choose the tail from H_1 , not from whichever tail gives a small number: $H_1: p > 0.1$ needs $P(X \geq \text{observed})$, an upper tail. Two-tailed tests double the tail when the claim is “different”, not “greater”.
- The hypotheses are about the PARAMETER (the true defect rate), not the sample; and they are fixed before the data are seen, otherwise the test is rigged.

■ Work from known quantities

- State the known values and their units. Choose the relation because its assumptions fit this case, then rearrange before substitution.
- Known: $H_0: p = 0.1$, $H_1: p > 0.1$, $n = 10$, and 3 defects observed. Under H_0 , $X \sim \text{binomial}(10, 0.1)$: $P(X \geq 3) = 1 - (0.9^{10} + 10 \times 0.1 \times 0.9^9 + 45 \times 0.01 \times 0.9^8) \approx 1 - 0.9298 = 0.0702$. At the 5% level, $0.0702 > 0.05$: insufficient evidence that the defect rate exceeds 10%.

Example:

A correct one-tailed p-value is 0.03 at the 5% level. State the conclusion using the same sequence: compare with the level $\rightarrow$ conclude in context. (Answer: $0.03 < 0.05$, so reject H_0 : the evidence supports the alternative claim at the 5% level.)

■ Check the conclusion and its limits

- Failing to reject H_0 is not proof it is true – the sample may simply be too small to detect a real difference. A 5% level means 1 in 20 innocent claims will be “rejected” by chance over many tests; significance does not measure importance.
- Return to the original observation. Explain what the result supports, which conditions it assumes, and one way to test a competing explanation.

Warn:

Failing to reject a null hypothesis proves it is true. This claim is false: the test only measures whether the evidence against H_0 clears the threshold. Weak evidence leaves H_0 unproven, not proven – “insufficient evidence” and “proven true” are entirely different conclusions.

Key:

Statistical hypothesis testing: state H_0 and H_1 in parameters first, compute the tail probability under H_0 , compare with the pre-set significance level, and conclude in context – never claim the null is proven when it survives.