The language of hypothesis testing

S5.1 · Statistics · revision notes

The language and logic of hypothesis testing: null and alternative hypotheses, significance level, test statistic, one- and two-tailed tests, critical value, critical region, acceptance region, p-value and actual significance level. Every hypothesis test on Paper 3 uses these ideas, and some parts ask you to define or explain them.

Key facts

- H_0: parameter = value; H_1: $<$ or $>$ (one-tailed), $\neq$ (two-tailed)
- Critical region: the values of the test statistic that lead to rejecting H_0
- p-value: probability of the observed value or a more extreme one, if H_0 is true
- Actual significance level $= P(\text{in the critical region} \mid H_0 \text{ true})$
- Test statistic: the value from the sample used to make the decision
- Critical value: the boundary (first value) of the critical region
- Reject H_0 if the p-value is less than the significance level, or if the test statistic is in the critical region
- Two-tailed test: half the significance level in each tail

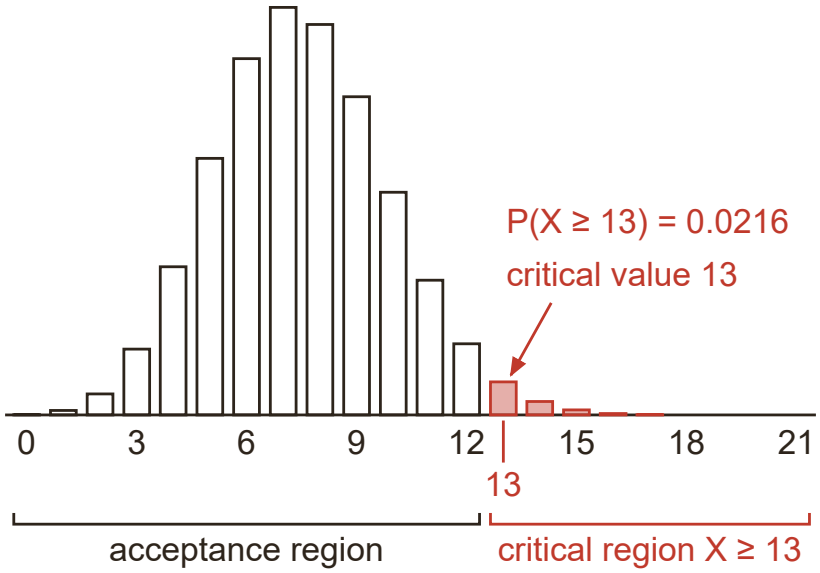
Notes

Hypotheses

- A **hypothesis test** uses a sample to test a claim about a population parameter: at AS a proportion p; at A level also a correlation coefficient ρ or a mean μ
- The **null hypothesis** H_0 gives the parameter a single value, e.g. $H_0 : p = 0.3$. It is assumed true unless there is enough evidence against it.
- The **alternative hypothesis** H_1 is what you suspect instead: $p > 0.3$ or $p < 0.3$ (**one-tailed**), or $p \neq 0.3$ (**two-tailed**).
- Hypotheses are always about the population parameter, never about the sample result.

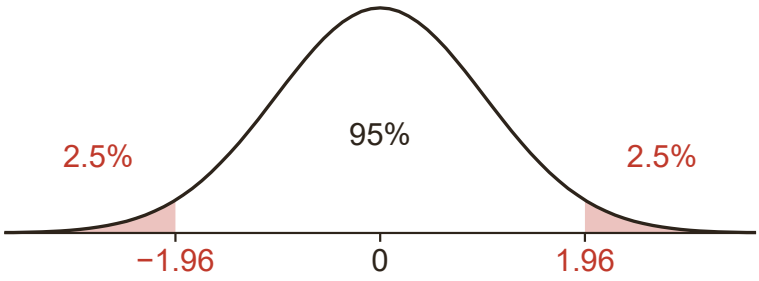
The key terms

- The **test statistic** is the value from the sample used in the test, e.g. the number of successes X
- The **significance level** (e.g. 5%) sets how unlikely a result must be before H_0 is rejected. It is also the probability of rejecting H_0 when H_0 is actually true (for a binomial test the true probability is at most this; see below).
- The **critical region** is the set of values of the test statistic that lead to rejecting H_0. The **critical value** is the first value in it (its boundary). The **acceptance region** is the rest.



H_0: p = 0.25, X ~ B(30, 0.25). Reject H_0 only if X lands in the red tail

- The **p-value** is the probability, if H_0 is true, of getting the observed value or a more extreme one. If the p-value is less than the significance level, reject H_0
- In a two-tailed test the significance level is split equally between the tails, e.g. 2.5% in each tail for a 5% test.



Two-tailed 5% test: 2.5% in each tail

Actual significance level

- With a discrete test statistic, the critical region usually cannot have a probability exactly equal to the significance level. The **actual significance level** is the probability of being in the critical region when H_0 is true.

Conclusions

- State the decision (reject or do not reject H_0), then a conclusion in context that does not claim certainty, e.g. ‘there is sufficient evidence to suggest that the proportion has increased’.
- Not rejecting H_0 does not prove that it is true; there is just not enough evidence against it.

How to answer each type of question

M = method mark, A = answer mark, B = a mark on its own.

Write suitable hypotheses

1 to 2 marks | Grade E

1. Identify the parameter and define it in words.
2. H_0: the parameter equals the claimed value.
3. H_1: choose $>$, $<$ or $\neq$ from the wording of the suspicion.

Example. A coin is thought to be biased towards heads. It is to be tossed 20 times to test this. Write down suitable hypotheses, defining your parameter.

Model answer. $H_0 : p = 0.5$ (B1), $H_1 : p > 0.5$ (B1), where p is the probability that the coin lands heads

Use a p-value to reach a conclusion

2 to 3 marks | Grade C

1. For a two-tailed test, compare the one-tail probability with half the significance level.
2. State the decision about H_0
3. Write a conclusion in context using ‘evidence’ or ‘suggest’.

Example. The proportion of a shop's customers who pay with cash has been 0.4. A manager tests $H_0 : p = 0.4$ against $H_1 : p \neq 0.4$ at the 5% level. For the number x in the sample who pay with cash, $P(X \geq x) = 0.034$ if H_0 is true. State the conclusion of the test.

Model answer.

The test is two-tailed, so compare with 0.025: $0.034 > 0.025$ (M1), so do not reject H_0 (A1)

There is insufficient evidence to suggest that the proportion of customers who pay with cash has changed (A1)

Critical region, actual significance level and what they mean

3 to 4 marks | **Grade B**

1. Define terms in words, linked to the test statistic.
2. Actual significance level: find the probability of the critical region using the distribution under H_0
3. If asked why it is not exactly the significance level, refer to the test statistic being discrete.

Example. A test of $H_0 : p = 0.25$ against $H_1 : p > 0.25$ uses $X \sim B(30, p)$. At the 5% level the critical region is $X \geq 13$.
(a) Explain what is meant by the critical region.
(b) Find the actual significance level of the test.
(c) Explain why the actual significance level is not 5%.

Model answer.

(a) The set of values of X for which H_0 would be rejected (B1)

(b) $P(X \geq 13) = 1 - P(X \leq 12)$ with $X \sim B(30, 0.25)$ (M1) $= 0.0216$ (A1)

(c) X is discrete, so the tail probabilities go up in steps: $P(X \geq 12) = 0.0507$ is more than 5%, so the largest allowed critical region has probability 0.0216 (B1)

Don't lose marks

- Writing hypotheses about the sample (e.g. x or r) instead of the parameter.
- Writing conclusions that claim certainty, e.g. 'this proves the coin is biased'.
- Forgetting to halve the significance level in a two-tailed test.
- Confusing the critical value (a value of the test statistic) with a probability.

More tips

Memory tricks

- H_0 always has an equals sign, and both hypotheses use the parameter (p, μ or ρ).
- Two-tailed: halve the significance level for each tail (or double the one-tail probability and compare it with the full level).
- Conclusion template: 'There is (in)sufficient evidence at the 5% level to suggest that …', finished in the words of the question.

Exam technique

- Define the parameter in words, e.g. 'p is the proportion of customers who pay with cash'.
- Always end with a conclusion in context that refers back to the claim in the question.
- If asked for the actual significance level, give the probability of the critical region, not the stated 5%.

What each grade needs

What you need to be able to do, from the first marks up to the top grade.

Grade E **Write null and alternative hypotheses**

H_0 gives the parameter a single value; H_1 says it is greater, less or different.

Grade E

Decide whether a test is one- or two-tailed

'Increased' or 'decreased' gives a one-tailed test; 'changed' or 'different' gives a two-tailed test.

Grade D

Explain what the significance level means

It sets how unlikely a result must be to reject H_0, and is the probability of rejecting H_0 when it is true.

Grade C

Define critical region and critical value

The critical region is the set of values that lead to rejecting H_0; the critical value is its boundary.

Grade C

Use a p-value to reach a decision

Reject H_0 if the p-value is less than the significance level.

Grade B

Find the actual significance level of a test

Calculate the probability that the test statistic falls in the critical region when H_0 is true.

Grade A

Interpret test results and their limitations

Explain that not rejecting H_0 does not prove it, and why the actual and stated significance levels differ.

Free notes, Full packs, practice papers and mock exams at chhetriacademy.com. Written for this site; not taken from past papers. Stuck? 1-to-1 tutoring with a free 30-minute trial lesson.