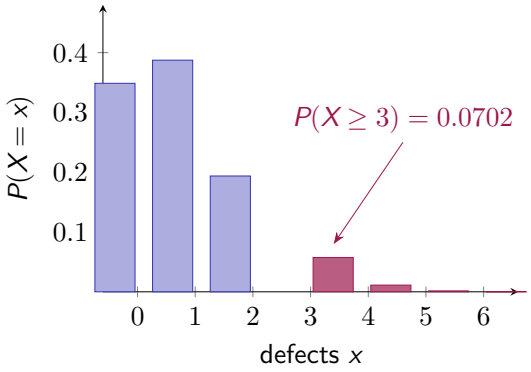
Statistical hypothesis testing

AQA A-level Mathematics 7357, section O: hypothesis tests for a binomial proportion, a normal mean and correlation

Every test has five steps

Step	**What you write**
1. Hypotheses	H_0 and H_1 about a parameter, before the data
2. Model	the distribution of the **test statistic** if H_0 is true
3. Probability	the tail at least as extreme as the result
4. Compare	with the **significance level**, say 5%
5. Conclude	in context, with no claim of proof

Is the defect rate above 10%?



- A factory claims a 10% defect rate.
- 3 of a sample of 10 are faulty.
- Is 3 surprising if the claim is true?

Hypotheses name a parameter

Correct

$H_0 : p = 0.1$

$H_1 : p > 0.1$

p is the true defect rate of the whole production.

Wrong

H_0: the sample has 1 defect.

$H_1 : X > 1$, about the sample.

Choosing H_1 after seeing the result.

Worked: the defect test

1. **Given** $H_0 : p = 0.1$, $H_1 : p > 0.1$, 5% level; $x = 3$ of 10
2. **Model** Under H_0, $X \sim B(10, 0.1)$
3. **Tail** $P(X \geq 3) = 1 - P(X \leq 2) = 1 - 0.9298 = 0.0702$
4. **Compare** $0.0702 > 0.05$
5. **Conclusion** Do not reject H_0: insufficient evidence that the defect rate exceeds 10%.

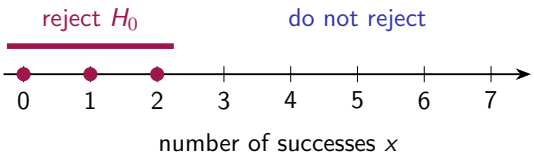
Decision

This does not prove the rate is 10%: a sample of 10 is small.

One tail or two?

Claim	H_1	**Tail**
the rate has increased	$p > 0.1$	upper
the rate has fallen	$p < 0.1$	lower
the rate has changed	$p \neq 0.1$	both, 2.5% each

A critical region, drawn



- $X \sim B(20, 0.3)$, $H_1 : p < 0.3$, 5%.
- $P(X \leq 2) = 0.0355 \leq 0.05$.
- $P(X \leq 3) = 0.1071 > 0.05$.
- **Critical region**: $X \leq 2$.

Two-tailed critical regions

1. **Given** $X \sim B(20, 0.3)$, $H_1 : p \neq 0.3$, 5% level
2. **Lower tail** $P(X \leq 1) = 0.0076$, but $P(X \leq 2) = 0.0355 > 0.025$: $X \leq 1$
3. **Upper tail** $P(X \geq 11) = 0.0171$, but $P(X \geq 10) = 0.0480 > 0.025$: $X \geq 11$
4. **Actual level** $0.0076 + 0.0171 = 0.0247$

Decision

Critical region $X \leq 1$ or $X \geq 11$; **actual significance level** 2.47%.

The actual significance level

For $X \leq 2$ under $B(20, 0.3)$ the actual level is 3.55%, below the nominal 5%.

- A count cannot hit 5% exactly.
- Actual level $= P(\text{in critical region} \mid H_0)$.
- It is the probability of rejecting a true H_0.

Actual significance level the probability of rejecting H_0 when H_0 is true.

Testing a normal mean

1. **Given** $\sigma = 4,\ n = 16,\ \bar{x} = 51.8;\ H_0 : \mu = 50,\ H_1 : \mu > 50$
2. **Model** Under H_0, $\overline{X} \sim N\left(50, \frac{4^2}{16}\right) = N(50, 1^2)$
3. **Statistic** $z = \frac{51.8-50}{1} = 1.8$
4. **Compare** $P(Z \geq 1.8) = 0.0359 < 0.05$, or $1.8 > 1.6449$
5. **Conclude** Reject H_0: evidence that the mean mass exceeds 50 g.

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Testing for correlation

1. **Hypotheses** $H_0 : \rho = 0$, $H_1 : \rho > 0$, 5% level
2. **Data** $n = 10$ pairs, $r = 0.62$
3. **Critical value** 0.5494 (one-tailed, 5%, $n = 10$)
4. **Conclude** $0.62 > 0.5494$: reject H_0; evidence of positive correlation.

Decision

Correlation is evidence of association, not of cause.

Wording the conclusion

Full marks

Reject H_0. There is evidence at the 5% level that the defect rate has increased.

Do not reject H_0. There is insufficient evidence that the mean has changed.

Loses the mark

"H_0 is true."

"This proves the rate has increased."

"Reject H_0." with no context.

Traps in hypothesis tests

Trap	**Fix**
Using $P(X = 3)$ as the evidence	use the tail, $P(X \geq 3)$
Tail chosen to look small	H_1 decides the tail
Two-tailed test at full 5% per tail	compare each tail with 2.5%
Not rejecting means H_0 is true	only "insufficient evidence"
Significant means important	significance is not size

Check yourself

Question	**Answer**
One-tailed **p-value** 0.03 at 5%?	reject H_0
"The coin is biased": H_1?	$p \neq 0.5$
$H_1 : p > 0.4$, observed 9: which probability?	$P(X \geq 9)$
Two-tailed 10% test: each tail?	5%
Distribution of $\overline{X}$ from $N(20, 9)$, $n = 4$?	$N(20, \frac{9}{4})$