Cambridge AS & A Level · Mathematics 9709 · Notes

Hypothesis tests A2

Read online at **pastopic.com/cambridge/mathematics-9709/notes/hypothesis-tests**

Last checked 29 September 2026

Read first: **Discrete random variables**, **The normal distribution**, **The Poisson distribution**, **Sampling and estimation**

What you need to know

By the end of this topic you should be able to:

1. **Explain what a hypothesis test is** and use its words correctly: null hypothesis, alternative hypothesis, significance level, rejection region (also called the critical region), acceptance region and test statistic. Know the difference between a one-tailed and a two-tailed test, and give every conclusion in the context of the question.
2. **Test a claim about a binomial or Poisson distribution from one observation** (one count, such as "4 out of 30" or "7 calls in 20 minutes"), either by working out the exact probabilities or, when it is suitable, with a normal approximation.
3. **Test a claim about a population mean** when the population is normal with a known variance, or when the sample is large (so the Central Limit Theorem applies and the variance can be estimated from the sample).
4. **Explain Type I and Type II errors** for a particular test, in context.
5. **Calculate the probabilities of Type I and Type II errors** for tests that use a normal distribution, or exact binomial or Poisson probabilities.

Tests with the t-distribution or χ^2, and tests comparing two samples, are not in this syllabus (they are Further Mathematics).

Understand it

The idea of a test

Someone makes a claim about a population, such as "35% of visitors click the advert" or "the mean mass is 1000 g". You suspect it is wrong, so you collect some data. A test asks one question:

If the claim were true, how likely is a result as extreme as the one we got?

If the answer is "very unlikely", the data are evidence against the claim, and you reject it. If the result could easily happen by chance, you have no real evidence against the claim, so you keep it. Keeping it does **not** prove it is true; you simply haven't shown it to be false.

The words

- **Null hypothesis,** H_0: the claim being tested, written as the parameter **equal** to one value, such as $H_0 : p = 0.35$, $H_0 : \lambda = 3$ or $H_0 : \mu = 1000$. The whole test is worked out assuming H_0 is true.
- **Alternative hypothesis,** H_1: what you suspect instead. $H_1 : p < 0.35$ or $H_1 : \mu > 1000$ is **one-tailed** (you are looking for a change in one direction only); $H_1 : \mu \neq 1000$ is **two-tailed** (you are looking for a change either way). The wording decides it: "decreased", "less than", "improved" or "more likely" means one tail; "changed", "different from" or "biased" (with no direction) means two.
- **Test statistic**: the number you calculate from the sample to decide, such as the count X or the value z from a sample mean.
- **Significance level,** α: how unlikely the result must be before you reject H_0, such as 5%. It is chosen before the data are seen.
- **Rejection (critical) region**: the values of the test statistic that lead you to reject H_0. The boundary value is the **critical value**. The other values form the **acceptance region**.

The parameter always belongs to the **population**, so write p, λ or μ (or "population mean"), never $\bar{x}$ or just "mean".

A binomial test

A website says 35% of visitors click an advert. A designer thinks a new layout has lowered this. She records 20 visitors, and X of them click. If the claim is true, $X \sim B(20, 0.35)$.

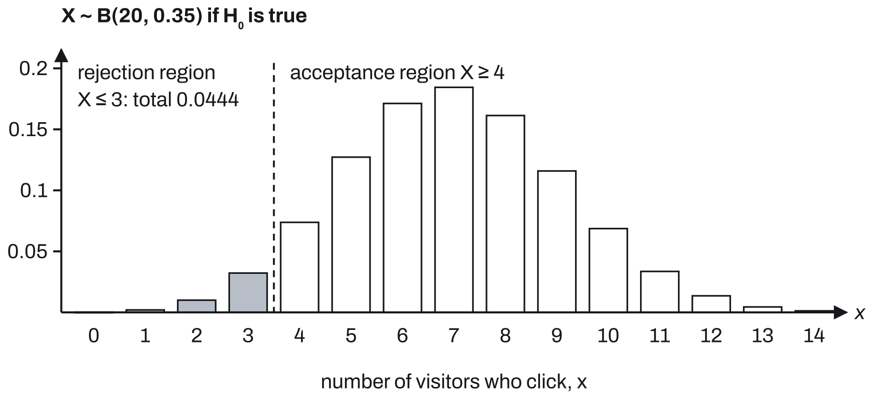
$H_0 : p = 0.35$, $H_1 : p < 0.35$ (one-tailed), at the 5% level.

Small values of X point towards H_1, so the evidence is measured by a **lower tail** probability. Suppose 3 visitors click. Then

$$P(X \leqslant 3) = 0.65^{20} + 20(0.65)^{19}(0.35) + \binom{20}{2}(0.65)^{18}(0.35)^2 + \binom{20}{3}(0.65)^{17}(0.35)^3 = 0.0444.$$

$0.0444 < 0.05$, so the result is in the rejection region: reject H_0. There is evidence that the proportion of visitors who click has decreased.

The rejection region itself is found by adding one more value at a time until the tail passes 5%: $P(X \leqslant 3) = 0.0444 < 0.05$ but $P(X \leqslant 4) = 0.1182 > 0.05$. So the rejection region is $X \leqslant 3$.



Two things to notice:

- You always use the **tail**, $P(X \leqslant 3)$, never the single value $P(X = 3)$. The question is "a result **this extreme or more**".
- A count can only take whole values, so the actual probability of landing in the rejection region is 0.0444, a little under 5%. For a binomial or Poisson test, this is the true size of the test.

A Poisson test

A Poisson test works the same way with $X \sim \mathrm{Po}(\lambda)$. First **scale the mean to the period observed**: if messages arrive at 10 per hour and you count for 12 minutes, $\lambda = 10 \times \frac{12}{60} = 2$. For $H_1 : \lambda > 2$, large counts are the evidence, so find an **upper** tail: after 6 messages, $P(X \geqslant 6) = 1 - P(X \leqslant 5)$. Leave the observed value out of the part you subtract.

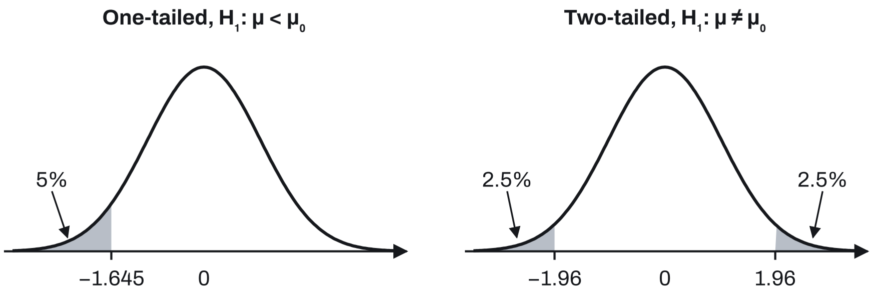
A test for a population mean

Flour bags are supposed to have mean mass 1000 g, with standard deviation 12 g. An inspector thinks the mean has fallen. He weighs a random sample of 36 bags and gets $\bar{x} = 996.2$ g.

If $H_0 : \mu = 1000$ is true, the sample mean has distribution $\bar{X} \sim N\left(1000, \frac{12^2}{36}\right)$, with standard deviation $\frac{12}{\sqrt{36}} = 2$. The test statistic is

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{996.2 - 1000}{2} = -1.9.$$

At the 5% level with $H_1 : \mu < 1000$, the critical value is $z = -1.645$ (from the critical values table, $P(Z \leqslant 1.645) = 0.95$). $-1.9 < -1.645$, so reject H_0: there is evidence that the mean mass has decreased.



shaded: rejection region for z

For a **two-tailed** test the significance level is split between the two ends, so at 5% the critical values are ± 1.96, not ± 1.645. The common critical values are:

Significance level	One-tailed	Two-tailed
10%	1.282	1.645
5%	1.645	1.960
2%	2.054	2.326
1%	2.326	2.576

You can compare in two ways; both are fine:

- **Compare z with the critical value**, as above.
- **Compare the tail probability with α**: $P(Z < -1.9) = 1 - \Phi(1.9) = 0.0287 < 0.05$. For a two-tailed test, compare the tail with $\frac{\alpha}{2}$.

When may you use the normal distribution for $\bar{X}$?

- The population is **normal** and σ is known: $\bar{X}$ is exactly normal for any n. The Central Limit Theorem is not needed.
- The population distribution is **unknown** but the sample is **large**: by the Central Limit Theorem, $\bar{X}$ is approximately normal. If σ is unknown, use the unbiased estimate s^2 from the sample in its place.

If the test uses an old value of σ (from "in the past"), you are assuming that the standard deviation has not changed. Questions that say "stating a necessary assumption" want this sentence.

Tests with a normal approximation

When n is large, a binomial count $X \sim B(n, p)$ is approximately $N\big(np,\ np(1-p)\big)$ (use it when $np > 5$ and $n(1-p) > 5$). A Poisson count $X \sim \mathrm{Po}(\lambda)$ with $\lambda > 15$ is approximately $N(\lambda, \lambda)$. A count is whole numbers, so use a **continuity correction**: $P(X \geqslant 28)$ becomes $P(Y > 27.5)$ and $P(X \leqslant 49)$ becomes $P(Y < 49.5)$.

For example, with $H_0 : \lambda = 20$, $H_1 : \lambda > 20$ and 28 events observed: $z = \frac{27.5-20}{\sqrt{20}} = 1.677 > 1.645$, so at the 5% level there is evidence that the mean has increased.

A proportion test from a large sample works the same way: the sample proportion is approximately $N\left(p, \frac{p(1-p)}{n}\right)$, and the question may just give you the value of z to compare.

Type I and Type II errors

A test can reach the wrong conclusion in two ways:

	H_0 **is really true**	H_0 **is really false**
Reject H_0	**Type I error**	correct
Do not reject H_0	correct	**Type II error**

- A **Type I error** is rejecting H_0 when it is true: concluding that the mean has decreased when it hasn't.
- A **Type II error** is not rejecting H_0 when it is false: finding no evidence of a decrease when the mean really has decreased.

So once you know the test's conclusion, only one error is possible: if H_0 was rejected, only a Type I error could have been made; if not, only a Type II.

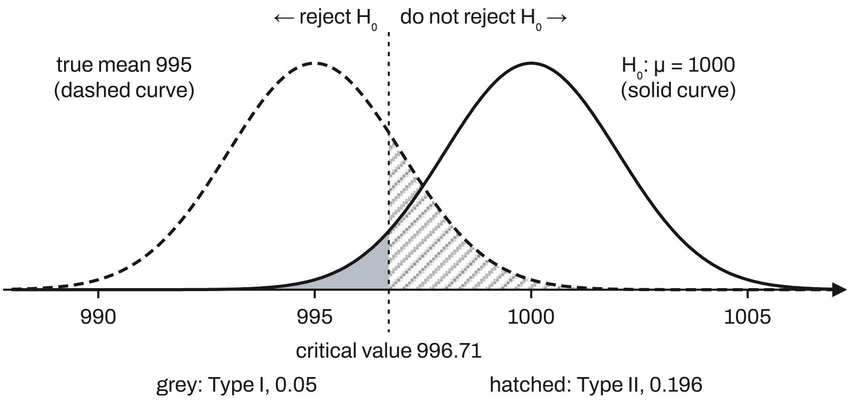
$P(\text{Type I error}) = P(\text{result in the rejection region, assuming } H_0)$.

- For a normal test this is exactly the significance level, such as 0.05.
- For a binomial or Poisson test it is the actual tail probability of the rejection region, such as 0.0444 in the advert test above, which is at most α.

$P(\text{Type II error}) = P(\text{result in the acceptance region, using the true value})$. You can only find it when you are told the true value of the parameter.

For the flour bags, reject H_0 when $\bar{x} < 1000 - 1.645 \times 2 = 996.71$. If the true mean is really 995, then $\bar{X} \sim N(995, 2^2)$, and a Type II error happens when $\bar{x} \geqslant 996.71$:

$$P(\text{Type II}) = P(\bar{X} > 996.71) = 1 - \Phi\left(\frac{996.71 - 995}{2}\right) = 1 - \Phi(0.855) = 0.196.$$



The diagram shows the trade-off: moving the critical value to the left makes the Type I area smaller but the Type II area bigger.

Writing the conclusion

A test gives **evidence**, not proof. So the conclusion must:

- say whether H_0 is rejected, with the comparison that shows why;
- be **in context** (the mass of flour bags, the proportion of visitors);
- **not be definite**: "there is (in)sufficient evidence that the mean mass has decreased", not "the mean mass has decreased";
- not contradict itself, such as "reject H_0; there is no evidence of a decrease".

Key facts and formulas

"Given" means it is in the MF19 list of formulae and tables in the exam. "Learn it" means you must remember it.

Formula	Status
Binomial $B(n,p)$: $P(X=r)=\binom{n}{r}p^r(1-p)^{n-r}$, $\mu=np$, $\sigma^2=np(1-p)$	Given
Poisson $\mathrm{Po}(\lambda)$: $P(X=r)=e^{-\lambda}\dfrac{\lambda^r}{r!}$, $\mu=\sigma^2=\lambda$	Given
Central Limit Theorem: $\bar{X}\sim N\left(\mu,\dfrac{\sigma^2}{n}\right)$	Given
Sample proportion: approximately $N\left(p,\dfrac{p(1-p)}{n}\right)$	Given
Unbiased estimates: $\bar{x}=\dfrac{\Sigma x}{n}$, $s^2=\dfrac{1}{n-1}\left(\Sigma x^2-\dfrac{(\Sigma x)^2}{n}\right)$	Given
Critical values table: $P(Z\leqslant z)=p$, e.g. $0.95\to1.645$, $0.975\to1.960$, $0.98\to2.054$, $0.99\to2.326$, $0.995\to2.576$	Given
Test statistic for a mean: $z=\dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$	Learn it
Critical value of $\bar{x}$: $\mu_0-z\dfrac{\sigma}{\sqrt{n}}$ (lower tail) or $\mu_0+z\dfrac{\sigma}{\sqrt{n}}$ (upper tail)	Learn it
Two-tailed test: compare each tail with $\frac{\alpha}{2}$ (e.g. ±1.96 at 5%)	Learn it
Normal approximations: $B(n,p)\approx N\big(np,np(1-p)\big)$; $\mathrm{Po}(\lambda)\approx N(\lambda,\lambda)$ for $\lambda>15$; with a continuity correction	Learn it
$P(\text{Type I})=P(\text{reject } H_0 \mid H_0 \text{ true}):=\alpha$ for a normal test, the rejection region's probability for a discrete one	Learn it
$P(\text{Type II})=P(\text{do not reject } H_0 \mid \text{the true value})$	Learn it

How to do it

In the Paper 6 questions we have tagged for this topic (71 questions, 2021–2025; 61 different, because some papers share questions), the types came up this often. 32 of the 61 mix two types, so the counts add up to more than 61.

Question type	Questions
Type I and Type II errors	32
Test for a population mean with the normal distribution	30
Binomial test with exact probabilities	19
Poisson test with exact probabilities	9
Test with a normal approximation to the binomial	3

1. Type I and Type II errors (32 questions)

Explaining an error in context. Say what the test **concludes** and what is **really** true, both in context:

- Type I: "concluding that the mean waiting time has decreased when in fact it has not".
- Type II: "concluding that there is no evidence that the proportion has fallen when in fact it has".

"Which error might have been made?" Look at the conclusion of the test:

- H_0 rejected: a Type I error is possible; a Type II error is not.
- H_0 not rejected: a Type II error is possible; a Type I error is not.

Give the reason ("because H_0 was rejected") every time, and answer the whole question: if it asks about both errors, say which one is possible **and** which one is not.

Probability of a Type I error.

1. **Normal test:** it is the significance level. At 2%, $P(\text{Type I}) = 0.02$.
2. **Binomial or Poisson test:** find the rejection region under H_0 (see types 3 and 4), then $P(\text{Type I}) = P(X \text{ in the rejection region})$. Show **both** tail probabilities, the one inside α and the next one past it, with a comparison each time, so the examiner can see the region is right.
3. **A rule is given** ("she will reject the claim if 2 or fewer ..."): $P(\text{Type I})$ is the probability of that rule's region under H_0, and this is also the test's actual significance level.
4. **Sample mean with a given cut-off** ("if the sample mean is below a set value she will conclude ..."): $P(\text{Type I}) = P(\bar{X} < \text{cut-off})$ with $\bar{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right)$.

Probability of a Type II error.

1. Find the rejection region **under** H_0, as a count or as a critical value of $\bar{x}$: $\mu_0 \pm z\frac{\sigma}{\sqrt{n}}$.
2. Switch to the **true** value of the parameter that the question gives.

3. $P(\text{Type II}) = P(\text{acceptance region})$ with the true value. For a count this is usually $1 - P(\text{rejection region})$; for a mean, standardise the critical value with the true mean and the same $\frac{\sigma}{\sqrt{n}}$.

For a Poisson test, scale the true rate to the period observed too.

Variations you will meet:

- **"State the probability of a Type I error"** for a normal test, with no working: just the significance level.
- **Only the true value is missing.** "What further information is needed to find $P(\text{Type II})$?" The true value of the parameter (in context, such as "the actual mean number of calls per hour").
- **Working backwards from $P(\text{Type II})$.** If the rule is "reject if there are no successes in 8 trials", then $P(\text{Type II}) = 1 - (1-p)^8$. Given it equals 0.75: $(1-p)^8 = 0.25$, so $p = 1 - 0.25^{1/8} = 0.159$.
- **A Type I error was made, so what was the level?** A Type I error means H_0 was rejected. If the result is significant at 5% but not at 1%, the test cannot have been at 1%.
- **Comment on a claim that an error was made.** If the claim is a Type II error but H_0 was rejected, the claim is impossible; say so, with the reason.

2. Test for a population mean with the normal distribution (30 questions)

1. **Hypotheses:** $H_0 : \mu = \mu_0$ and $H_1 : \mu < \mu_0, \mu > \mu_0$ or $\mu \neq \mu_0$, from the wording. Write μ or "population mean".
2. **Find $\bar{x}$, and σ or s.** From totals: $\bar{x} = \frac{\Sigma x}{n}$ and $s^2 = \frac{1}{n-1}\left(\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right)$. Keep 4 or more significant figures.
3. **Standardise:** $z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$. The $\sqrt{n}$ must be there.
4. **Compare** with the critical value for the level and the number of tails (or compare the tail probability with α, or $\frac{\alpha}{2}$ for two tails). Write the comparison out, such as "$-2.21 < -2.054$".
5. **Conclude in context**, not definitely.

Variations you will meet:

- **"Stating a necessary assumption"**: the standard deviation (or variance) is unchanged from the past value.
- **"Why a one-tailed (or two-tailed) test?"**: quote the direction the person is looking for, such as "she wants to know if the mean has **increased**, not just changed", or "he is looking for a difference in either direction".
- **The value of z is given.** Just compare it with the right critical value and conclude.
- **The critical value method.** Find the boundary for $\bar{x}$, such as $\mu_0 - 2.054\frac{\sigma}{\sqrt{n}}$, and compare $\bar{x}$ with it.
- **Set of values of $\bar{x}$ that lead to rejection** (or the smallest or largest one): set $\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ equal to the critical value, solve for $\bar{x}$, then give the inequality in the right direction.
- **An unknown in the formula** (σ, n, or μ_0), when you are told the value of z or that the result "just leads to rejection": set z equal to the given value or the critical value and solve. For example, if a sample of 64 with $\sigma = 4$ gives $\bar{x} = 31.2$ and this just leads to rejection of $H_0 : \mu = m$ in favour of $\mu < m$ at 5%, then $\frac{31.2-m}{0.5} = -1.645$ and $m = 32.02$ (to 4 s.f.).

- **Set of significance levels.** Find the tail probability for the z-value (for example $z = -2.17$ gives $1 - \Phi(2.17) = 0.0150$). The result is significant at every level $\alpha\%$ above this, so $\alpha > 1.50$. For a two-tailed test, double the tail first.
- **"Significant at 5% but not at 1%"**: show the comparison at both levels, such as $1.645 < 2.1 < 2.326$.
- **The Central Limit Theorem.** Needed when the population distribution is not known (or not normal) and the sample is large; not needed when the population is normal. Say which, and why.
- **"Why use a sample?"**: testing destroys the item (a match can only be struck once), or the population is too large to test all of it. A sample from one group (one class, one day) is not representative of the whole population.

3. Binomial test with exact probabilities (19 questions)

1. **Hypotheses** about p: $H_0 : p = p_0$ and $H_1 : p < p_0$ (or $>$, or $\neq$). Use p or "proportion", not μ and not a count.
2. **Distribution under H_0:** $X \sim B(n, p_0)$.
3. **Tail probability** in the direction of H_1, including the observed value: $P(X \leqslant x)$ for "less", $P(X \geqslant x) = 1 - P(X \leqslant x - 1)$ for "more". Write the terms out, such as $0.65^{20} + 20(0.65)^{19}(0.35) + \ldots$
4. **Compare** with α ($\frac{\alpha}{2}$ for two tails) and **conclude** in context.

Variations you will meet:

- **Find the rejection region.** Work out the tail one value at a time until it passes α: the last value inside is the critical value. Show both the tail that is below α and the next one that is above it.
- **"The largest (or smallest) value of r in the rejection region"**: the same method; the answer is the critical value.
- **"Find the significance level of the test"** for a given rule: the probability of the rule's region under H_0.
- **Why a single probability is not enough.** A claim like "$P(X = 8) < 0.05$, so reject H_0" is wrong: the test needs the probability of the result **or anything more extreme**, $P(X \geqslant 8)$.
- **"Stating a necessary assumption"**: the trials are independent and the probability is the same for every trial, in context ("whether one seed grows does not affect whether another does").
- **"Why not use a Poisson approximation?"**: it needs n large ($n > 50$) and np small ($np < 5$); say which fails, with the numbers.
- **Comment on a model.** A model $B(n, p)$ can't give more than n successes; if more than n are possible in reality, the model is not suitable.

4. Poisson test with exact probabilities (9 questions)

1. **Scale** λ to the period in the question: $\mathrm{Po}(1.4)$ per day becomes $\mathrm{Po}(7)$ for 5 days.
2. **Hypotheses** with λ (or μ): $H_0 : \lambda = 7$, $H_1 : \lambda < 7$. Either the scaled value or the original rate is fine, if it is clear.
3. **Tail probability** in the direction of H_1, terms written out: $P(X \leqslant x) = e^{-\lambda}\left(1 + \lambda + \frac{\lambda^2}{2!} + \ldots\right)$, or $P(X \geqslant x) = 1 - P(X \leqslant x - 1)$.

4. **Compare** and **conclude** in context.

Variations you will meet:

- **Critical region and** $P(\text{Type I})$, exactly as for the binomial: show both tail probabilities either side of α.
- **Conclusion from the critical region.** Once you have it, a new count just needs "3 is in the critical region, so reject H_0".
- **Validity of the model.** A Poisson model needs events that happen independently, at a constant average rate. If the rate has been rising steadily over the years, the mean is not constant, so the test may not be valid.
- **A large mean.** If $\lambda > 15$, you may use $N(\lambda, \lambda)$ with a continuity correction instead, then standardise and compare as in type 5.

5. Test with a normal approximation to the binomial (3 questions)

1. **Hypotheses** about p, often two-tailed ("biased", "different from").
2. **Approximate:** $X \sim B(n, p_0) \approx N\big(np_0, np_0(1-p_0)\big)$.
3. **Continuity correction** in the direction of the tail, then standardise: for x below the mean, $z = \dfrac{x + 0.5 - np_0}{\sqrt{np_0(1-p_0)}}$.
4. **Compare** with the critical value (± 1.96 at 5% two-tailed, ± 2.326 at 2% two-tailed) and conclude.

Variations you will meet:

- **The value of z is given** for a proportion from a large sample. Explain the number of tails from the wording, write $H_0 : p = p_0$ and H_1 with the population proportion, compare z with the critical value, and conclude.

Worked examples

The bold codes show where marks are earned, as in Cambridge mark schemes. The number is how many marks.

- **M1** method mark: for a correct method, even if the numbers slip.
- **A1** accuracy mark: for a correct answer; it needs the M mark before it.
- **B1** mark for a correct result on its own, with no method needed.

Example 1

A call centre used to receive calls at a mean rate of 9 per hour. After an advertising campaign, the manager wants to test whether the mean rate has increased. In a randomly chosen 20-minute period, 7 calls are received. You may assume the number of calls has a Poisson distribution.

(a) Carry out the test at the 5% significance level. **[5]**

(b) State, with a reason, which type of error might have been made. **[2]**

(a) In 20 minutes the mean is $9 \times \frac{20}{60} = 3$. $H_0 : \lambda = 3$, $H_1 : \lambda > 3$. **B1**

$$P(X \geqslant 7) = 1 - e^{-3}\left(1 + 3 + \frac{3^2}{2!} + \frac{3^3}{3!} + \frac{3^4}{4!} + \frac{3^5}{5!} + \frac{3^6}{6!}\right) = 1 - 0.9665 = 0.0335.$$

M1 for $1 - P(X \leqslant 6)$ with $\lambda = 3$ and no end errors, **A1**.

$0.0335 < 0.05$, so reject H_0. **M1** There is evidence that the mean rate of calls has increased. **A1** (in context, not definite)

(b) H_0 was rejected, so a Type I error might have been made (concluding that the rate has increased when it has not). **B1 B1**

Example 2

A tennis player gets her first serve in with probability 0.45. Her coach teaches her a new technique and wants to test, at the 5% significance level, whether the probability has increased. She will take 16 first serves, and X of them go in.

(a) Find the rejection region for the test, and state the probability of a Type I error. **[5]**

(b) 10 of her serves go in. Carry out the test. **[2]**

(c) It is later found that the new technique has raised the probability to 0.75. Find the probability that the test makes a Type II error. **[3]**

(a) $H_0 : p = 0.45$, $H_1 : p > 0.45$, and under H_0, $X \sim B(16, 0.45)$. **B1**

$$P(X \geqslant 11) = \binom{16}{11}0.45^{11}0.55^5 + \binom{16}{12}0.45^{12}0.55^4 + \ldots + 0.45^{16} = 0.0486 < 0.05.$$

M1 A1

$$P(X \geqslant 10) = 0.0486 + \binom{16}{10}0.45^{10}0.55^6 = 0.0486 + 0.0755 = 0.124 > 0.05.$$

M1 for the next tail and a comparison. So the rejection region is $X \geqslant 11$, and $P(\text{Type I}) = 0.0486$. **A1**

(b) 10 is not in the rejection region ($P(X \geqslant 10) = 0.124 > 0.05$), so do not reject H_0. **M1** There is insufficient evidence that the probability of her first serve going in has increased. **A1**

(c) A Type II error happens when $X \leqslant 10$ although $p = 0.75$. **M1** for using $B(16, 0.75)$ with the acceptance region.

$$P(X \leqslant 10) = 1 - P(X \geqslant 11) = 1 - 0.810 = 0.190.$$

M1 for $1-$ the six terms from 11 to 16, **A1**.

Example 3

The lengths, x minutes, of a random sample of 80 phone calls to a help desk are summarised by

$$n = 80, \qquad \Sigma x = 544, \qquad \Sigma x^2 = 3876.95.$$

The help desk claims that the mean length of calls is 6.5 minutes.

(a) Test, at the 10% significance level, whether the population mean length is different from 6.5 minutes. **[8]**

(b) Find the set of significance levels, $\alpha\%$, at which this sample would lead to rejecting the claim. **[2]**

(c) Explain whether the Central Limit Theorem was needed in part (a). **[1]**

(a) $\bar{x} = \dfrac{544}{80} = 6.8$. **B1**

$$s^2 = \frac{1}{79}\left(3876.95 - \frac{544^2}{80}\right) = \frac{3876.95 - 3699.2}{79} = \frac{177.75}{79} = 2.25.$$

M1 A1

$H_0 : \mu = 6.5$, $H_1 : \mu \neq 6.5$, where μ is the population mean length. **B1**

$$z = \frac{6.8 - 6.5}{\sqrt{2.25/80}} = \frac{0.3}{0.16771} = 1.789.$$

M1 for standardising with $\sqrt{80}$, **A1**.

Two-tailed at 10%: $1.789 > 1.645$, so reject H_0. There is evidence that the population mean length of calls is not 6.5 minutes. **M1 A1**

(b) The tail is $1 - \Phi(1.789) = 1 - 0.9632 = 0.0368$. Two-tailed, so double it: 0.0736. **M1** The claim is rejected when $\alpha > 7.36$. **A1**

(c) Yes: the population distribution of call lengths is not known, so the Central Limit Theorem is needed to say that the sample mean is approximately normal (the sample is large). **B1**

Example 4

A machine fills bottles with a mean volume of 250 ml and standard deviation 4.8 ml. After a service, the manager tests whether the mean volume has increased, using a random sample of 40 bottles and a 1% significance level.

(a) The sample mean is 251.6 ml. Stating a necessary assumption, carry out the test. **[6]**

(b) Find the critical value of the sample mean for this test. **[2]**

(c) A similar test is later carried out on another sample of 40 bottles. Given that the mean volume is actually 252.5 ml, find the probability of a Type II error. **[3]**

(a) Assume the standard deviation is still 4.8 ml after the service. **B1**

$H_0 : \mu = 250$, $H_1 : \mu > 250$. **B1**

$$z = \frac{251.6 - 250}{4.8/\sqrt{40}} = \frac{1.6}{0.75895} = 2.108.$$

M1 A1

$2.108 < 2.326$, so do not reject H_0. **M1** There is insufficient evidence that the mean volume has increased. **A1**

(b) $250 + 2.326 \times \frac{4.8}{\sqrt{40}} = 250 + 1.765 = 251.77$ ml (to 5 s.f.). **M1 A1** H_0 is rejected when $\bar{x} > 251.77$.

(c) A Type II error is not rejecting H_0: $\bar{x} \leqslant 251.77$, when (using the unrounded critical value 251.765 below) $\bar{X} \sim N\left(252.5, \frac{4.8^2}{40}\right)$. **M1**

$$P(\bar{X} < 251.765) = \Phi\left(\frac{251.765 - 252.5}{0.75895}\right) = \Phi(-0.968) = 1 - 0.8335 = 0.167.$$

M1 for standardising the critical value with the true mean, **A1**.

Common mistakes

- **Conclusions that are definite or have no context.** Examiners mention this in almost every report. "The mean has decreased" or "reject H_0" alone lose the last mark. Write "there is (in)sufficient evidence that the mean time has decreased".
- **Vague hypotheses.** $H_0 = 0.3$, "mean $= 40$", or $\bar{x} = 40$ are not accepted. Write $H_0 : p = 0.3$ or $H_0 : \mu = 40$ (population mean). Use p for a proportion, λ for a Poisson mean and μ for a population mean, and write both H_0 and H_1.
- **The wrong number of tails.** Check the wording. A two-tailed test at 5% uses ± 1.96; using 1.645 (or 1.96 for a one-tailed test) costs marks.
- **Using a single probability instead of the tail.** $P(X = 6)$ or $P(X < 6)$ instead of $P(X \leqslant 6)$, or leaving the observed value out of an upper tail ($P(X \geqslant 7) = 1 - P(X \leqslant 6)$, not $1 - P(X \leqslant 7)$).
- **Not showing the working.** For binomial and Poisson tests, write the terms of the sum, such as $e^{-3}(1 + 3 + 4.5 + \ldots)$. A correct probability with no working usually earns only one mark.
- **Forgetting $\sqrt{n}$** when standardising a sample mean, or rounding values like s^2 too early.
- **Using the biased variance** (dividing by n) when estimating σ^2 from summary data.
- **A comparison that isn't shown.** Write "$0.0335 < 0.05$" or "$2.108 < 2.326$" before the conclusion; a conclusion with no comparison doesn't score.
- **Mixing up the errors.** "Rejecting H_0" and "H_0 being false" are different things. A Type I error is rejecting a **true** H_0; a Type II error is keeping a **false** one.
- **Type II errors with the wrong distribution.** Find the critical region with the H_0 value, then use the **true** value to find the probability of the acceptance region. Using the H_0 value twice, or finding the rejection region's probability instead of $1-$ it, are the usual slips.
- **Answering only half of "Type I, Type II or both?"** Say which error is possible **and** which is not, with the reason.

- **Not scaling the Poisson mean** to the period observed.

How to get the marks

- **Hypotheses:** one mark, for both H_0 and H_1 with the right symbol and value. It is often available even if the rest goes wrong, so always write them.
- **The calculation:** a method mark for the right tail (or standardising, with $\sqrt{n}$) and an accuracy mark for its value to 3 significant figures, from working that is shown.
- **The comparison:** a separate method mark. Compare like with like: a tail probability with α (or $\frac{\alpha}{2}$), or z with the critical value.
- **The conclusion:** the last mark is for a conclusion in context, not definite, and with no contradictions. It usually follows through from a wrong value.
- **Critical regions:** show the probability just inside **and** just outside the region, each compared with α. Then state the region, such as $X \leqslant 3$.
- **"State"** the probability of a Type I error in a normal test: just the significance level, no working needed.
- **"Stating a necessary assumption"** earns a mark on its own: the standard deviation is unchanged, or the trials are independent with a constant probability.
- **"Explain"** an error: the conclusion reached and the real situation, both in context.
- **"Use a binomial distribution"** means exact probabilities, not a normal or Poisson approximation.
- **"Show that"** the result is significant: write the comparison and the conclusion in full.
- **Accuracy:** 3 significant figures, from working kept to 4 or more; z to 3 decimal places is safest.

Quick check

1. Write down H_0 and H_1, and say whether the test is one-tailed or two-tailed. (a) A dentist says patients wait 12 minutes on average; a patient thinks the mean wait is longer. (b) A student wants to test whether a coin is biased.
2. In a one-tailed test of $H_0 : \mu = \mu_0$ against $H_1 : \mu < \mu_0$, the test statistic is $z = -1.93$. Find the set of significance levels, $\alpha\%$, at which H_0 is rejected.
3. A political party claims that 40% of voters support it. In a random sample of 150 voters, 49 support it. Use a normal approximation to test, at the 5% significance level, whether the support is less than 40%.
4. The number of faults in a 200 m roll of cable has a Poisson distribution, with mean 6.4 in the past. After a change in the process, a test at the 5% level of whether the mean has decreased uses one roll. Find the rejection region and the probability of a Type I error.
5. In question 3, which type of error might have been made? Describe it in context.

Answers

1. (a) $H_0 : \mu = 12$, $H_1 : \mu > 12$, where μ is the population mean waiting time; one-tailed. (b) $H_0 : p = 0.5$, $H_1 : p \neq 0.5$, where p is the probability of a head; two-tailed.
2. $1 - \Phi(1.93) = 1 - 0.9732 = 0.0268$. H_0 is rejected at every level above this, so $\alpha > 2.68$.
3. $H_0 : p = 0.4$, $H_1 : p < 0.4$. $X \sim B(150, 0.4) \approx N(60, 36)$. $P(X \leqslant 49) \approx \Phi\left(\frac{49.5-60}{6}\right) = \Phi(-1.75) = 0.0401$. $0.0401 < 0.05$ (or $-1.75 < -1.645$), so reject H_0: there is evidence that the support for the party is less than 40%.
4. $P(X \leqslant 2) = e^{-6.4}\left(1 + 6.4 + \frac{6.4^2}{2}\right) = e^{-6.4}(27.88) = 0.0463 < 0.05$. $P(X \leqslant 3) = 0.0463 + e^{-6.4}\frac{6.4^3}{6} = 0.0463 + 0.0726 = 0.119 > 0.05$. So the rejection region is $X \leqslant 2$, and $P(\text{Type I}) = 0.0463$.
5. H_0 was rejected, so a Type I error: concluding that the support is less than 40% when it is really 40%.

Past questions to try

- **9709/62/O/N/22 Q4**: a Poisson test with a Type I error, then a Type II error.
- **9709/62/O/N/21 Q6**: the rejection region of a binomial test and a Type II error.
- **9709/62/F/M/23 Q6**: reading a given z, a set of significance levels, and an unknown mean.
- **9709/62/M/J/21 Q1**: a two-tailed test with a normal approximation to the binomial.