IB Math AI HL — Topic 4

Hypothesis Testing for Population Parameters

One-Sample t-test · Two-Sample t-test · Binomial Test · Poisson Test · Correlation Test · Type I & II Errors

Difficulty	EASY	MEDIUM	HARD
Questions	10	10	10
Total Marks	30	40	50

Scan me for help

Name:______________________________ Date:_______________

Score:__________/120

Instructions: Answer all questions. Show all working including hypotheses, p-value or test statistic, comparison to significance level, and a contextual conclusion. Give answers correct to 3 significant figures unless stated. GDC required throughout.

Key Formulae

One-sample t-test statistic:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Two-sample t-test:

GDC: 2-SampTTest; assume equal variances unless stated

Binomial test: $X \sim B(n, p_0)$
Find $P(X \le k)$ or $P(X \ge k)$ for one-tailed;
Compare to α (or $\alpha/2$ for two-tailed)

Poisson test: $X \sim \text{Po}(\lambda_0)$
Find $P(X \le k)$ or $P(X \ge k)$
Compare to α (or $\alpha/2$ for two-tailed)

Correlation test (r vs r_{crit}):

$H_0 : \rho = 0$; $H_1 : \rho \ne 0$ (or $> 0, < 0$)

Percentage error:

$$\%\text{ error} = \frac{|v_A - v_E|}{|v_E|} \times 100$$

Type I error: $P(\text{reject } H_0 \mid H_0 \text{ true}) = \alpha$

Type II error: $P(\text{fail to reject } H_0 \mid H_0 \text{ false}) = \beta$

GDC Tips

- **One-sample t-test:** STAT $\to$ TESTS $\to$ T-Test. Enter μ_0, $\bar{x}$, s_x, n, and select the correct tail.
- **Two-sample t-test:** STAT $\to$ TESTS $\to$ 2-SampTTest. Enter both sample statistics. Select Pooled: No unless told equal variances.
- **Binomial p-value:** Use $\texttt{binomcdf}(n, p, k)$ for $P(X \le k)$; for upper tail compute $1 - \texttt{binomcdf}(n, p, k{-}1)$.
- **Poisson p-value:** Use $\texttt{poissoncdf}(\lambda, k)$ for $P(X \le k)$; for upper tail compute $1 - \texttt{poissoncdf}(\lambda, k{-}1)$.

- **Correlation p-value:** Enter data in lists $\rightarrow$ STAT $\rightarrow$ CALC $\rightarrow$ LinReg(a+bx) $\rightarrow$ DiagnosticOn to get r; compare $|r|$ to critical value table, or use 2-variable test.
- **Critical regions:** For discrete tests, find the smallest/largest k where the tail probability $\leq \alpha$.

Section A — Easy EASY (10 questions × 3 marks = 30 marks)

Scan me for help

1. EASY Calculator **[3 marks]**

A nutritionist claims that a particular brand of energy bar contains a mean of 250 kJ per bar. A random sample of 18 bars is tested and found to have a mean energy content of $\bar{x} = 244$ kJ with a sample standard deviation of $s = 12$ kJ.

Test the nutritionist's claim at the 5% significance level.

(a) State suitable hypotheses for this test. [1]

(b) Write down the p-value for this test. [1]

(c) State the conclusion of the test, giving a reason. [1]

2. EASY Calculator **[3 marks]**

A manufacturer claims that 20% of items produced by a machine are defective. A quality inspector randomly selects 15 items and finds 6 defective items.

Test, at the 10% significance level, whether the proportion of defective items is greater than 20%.

(a) State suitable hypotheses for this test, defining the parameter used. [1]

(b) Find the p-value for this test. [1]

(c) State the conclusion of the test, giving a reason. [1]

3. EASY Calculator **[3 marks]**

A call centre receives a mean of 4 complaints per hour under normal operating conditions. On a particular day, 9 complaints are received in a single hour.

Test, at the 5% significance level, whether the mean rate of complaints has increased.

(a) State suitable hypotheses for this test. [1]

(b) Find the p-value for this test. [1]

(c) State the conclusion of the test, giving a reason. [1]

4. EASY Calculator **[3 marks]**

A researcher investigates whether altitude and mean daytime temperature are linearly correlated. Data from 12 weather stations gives a Pearson correlation coefficient of $r = -0.743$.

The critical value of r for a two-tailed test at the 5% significance level with $n = 12$ is ± 0.576.

(a) State suitable hypotheses for this test. [1]

(b) State whether the result is significant, giving a reason. [1]

(c) Write a conclusion in context. [1]

5. EASY Calculator **[3 marks]**

A hypothesis test is conducted at the 1% significance level.

(a) State what is meant by a Type I error. [1]

(b) State the probability of a Type I error for this test. [1]

(c) A hypothesis test gives a p-value of 0.008. State the conclusion of the test. [1]

6. EASY Calculator **[3 marks]**

A ceramics factory states that the mean breaking strength of its tiles is $\mu = 380$ N. A sample of 10 tiles is tested, giving $\bar{x} = 374$ N and $s = 14$ N.

Assuming the breaking strengths are normally distributed, test at the 10% significance level whether the mean breaking strength differs from 380 N.

(a) State suitable hypotheses. [1]

(b) Write down the p-value. [1]

(c) State the conclusion in context. [1]

7. EASY Calculator **[3 marks]**

A seed supplier claims that 75% of seeds of a certain variety germinate. A gardener plants 20 seeds and observes that only 11 germinate.

Test at the 5% significance level whether the germination rate is less than 75%.

(a) State suitable hypotheses. [1]

(b) Find the p-value. [1]

(c) State the conclusion in context. [1]

8. EASY Calculator **[3 marks]**

(a) State what is meant by a Type II error in hypothesis testing. [1]

(b) In a hypothesis test with $H_0 : \mu = 50$, a Type II error occurs. State what this means in terms of H_0 and the true population mean. [1]

(c) Explain one way a researcher can reduce the probability of a Type II error while keeping the significance level fixed. [1]

9. EASY Calculator **[3 marks]**

Accidents at a road junction occur at a mean rate of 3 per month. Following the installation of new road markings, only 0 accidents are recorded in the next month.

Test at the 5% significance level whether the mean monthly accident rate has decreased.

(a) State suitable hypotheses for this test. [1]

(b) Find the p-value. [1]

(c) State the conclusion in context. [1]

10. EASY Calculator **[3 marks]**

The table below shows the number of hours studied per week and the score achieved in a test for a group of 8 students.

Hours studied	2	4	5	6	7	9	11	13
Test score	38	45	52	57	61	68	74	81

A Pearson correlation coefficient of $r = 0.997$ is found for this data. The critical value for a one-tailed test at the 1% level with $n = 8$ is 0.748.

(a) State suitable hypotheses for testing whether hours studied and test score are positively correlated. [1]

(b) Determine whether the result is significant at the 1% level. [1]

(c) Write a conclusion in context. [1]

Section B — Medium MEDIUM (10 questions × 4 marks = 40 marks)

Scan me for help

11. MEDIUM Calculator **[4 marks]**

A dairy company claims that its cartons of milk contain a mean volume of 1000 ml. A consumer group measures a random sample of 25 cartons and records the following results: $\bar{x} = 993$ ml, $s = 18$ ml.

(a) State the hypotheses for a two-tailed test of the company's claim. [1]

(b) State an assumption required for this test to be valid. [1]

(c) Find the p-value for the test. [1]

(d) State the conclusion at the 5% significance level, in context. [1]

12. MEDIUM Calculator **[4 marks]**

A biologist claims that a particular species of frog has a 0.35 probability of being female. To test this claim, a random sample of 20 frogs is caught.

Let X be the number of female frogs in the sample.

(a) State suitable hypotheses, defining the parameter used. [1]

(b) Find the critical region for a two-tailed test at the 5% significance level. [2]

(c) State the actual significance level of this test. [1]

13. MEDIUM Calculator **[4 marks]**

Meteors are observed at a remote observatory at a mean rate of 2.5 per night. An astronomer wants to test whether the rate has changed on a night when an unusual number of meteors is observed.

Let X be the number of meteors observed in a single night.

(a) State suitable hypotheses for a two-tailed test. [1]

(b) Find the critical region for a two-tailed test at the 10% significance level. [2]

(c) State the probability of a Type I error for this test. [1]

14. MEDIUM Calculator **[4 marks]**

A fitness coach compares the reaction times (in milliseconds) of two groups of athletes. The table gives the sample statistics.

Group	Sample size	Mean (ms)	Standard deviation (ms)
Sprinters	14	182	23
Swimmers	16	197	27

Test at the 5% significance level whether there is a difference in mean reaction time between the two groups. Assume both populations are normally distributed with equal variances.

(a) State suitable hypotheses. [1]

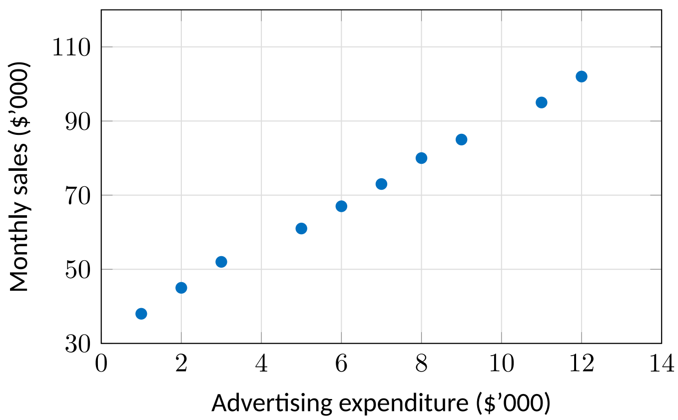
(b) Write down the p-value for this test. [1]

(c) State the conclusion in context. [1]

(d) State one assumption already stated in the question that is required for the test. [1]

15. MEDIUM Calculator **[4 marks]**

The scatter diagram below shows the advertising expenditure (in thousands of dollars) and monthly sales revenue (in thousands of dollars) for 10 branches of a retail chain.



The Pearson correlation coefficient for this data is $r = 0.9983$. The critical value for a one-tailed test at the 1% significance level with $n = 10$ is 0.715.

(a) State suitable hypotheses to test whether higher advertising expenditure is associated with higher monthly sales. [1]

(b) Determine whether the correlation is significant at the 1% level. Justify your answer. [2]

(c) Give one reason why the scatter diagram supports the use of a linear model. [1]

16. MEDIUM Calculator **[4 marks]**

A delivery service claims that 90% of parcels arrive on time. A logistics manager tests this claim using a sample of 30 parcels, with a one-tailed test at the 5% significance level with $H_0 : p = 0.9$, $H_1 : p < 0.9$.

The critical region is $X \leq 23$, where X is the number of parcels that arrive on time.

(a) Verify that $X \leq 23$ is the correct critical region. [1]

(b) Find $P(\text{Type I error})$ for this test. [1]

(c) Find $P(\text{Type II error})$ if the true probability of on-time delivery is 0.80. [2]

17. MEDIUM Calculator **[4 marks]**

A packaging machine fills bags of rice to a claimed mean mass of $\mu_0 = 500$ g. A factory inspector conducts a one-sample t-test using a random sample of n bags. She finds $\bar{x} = 495$ g and $s = 8.4$ g, and obtains a p-value of 0.0310.

(a) State the hypotheses for a two-tailed test of the machine's claim. [1]

(b) State the conclusion of the test at the 5% significance level, with a reason. [1]

(c) Given that the t-statistic is $t = -2.38$ (to 3 s.f.), find the value of n. [2]

18. MEDIUM Calculator **[4 marks]**

During regular production, a machine produces a mean of 1.8 faulty components per hour. After servicing, the machine produces 5 faulty components in a 2-hour period.

Test, at the 5% significance level, whether the mean rate of faults has changed.

(a) State suitable hypotheses for this two-tailed test, in terms of the mean number of faults per **hour**. [1]

(b) Explain how to adapt the Poisson parameter for observing over a 2-hour period. [1]

(c) Find the p-value for this test. [1]

(d) State the conclusion in context. [1]

19. MEDIUM Calculator **[4 marks]**

A pharmacologist measures the concentration of a drug (in μg/mL) in the bloodstream of patients receiving two different formulations.

The table gives the sample data.

Formulation	**Sample size**	**Mean (μg/mL)**	**Std dev (μg/mL)**
Tablet	20	3.42	0.68
Capsule	18	3.81	0.74

(a) Write down the hypotheses for a two-tailed test of whether the two formulations give the same mean concentration. [1]

(b) Write down the p-value for this test. (Do not assume equal variances.) [1]

(c) State the conclusion at the 5% significance level, in context. [1]

(d) State one assumption required for the validity of this test. [1]

20. MEDIUM Calculator **[4 marks]**

Eight candidates in a regional cooking competition are ranked by two judges. The rankings are given in the table below.

Candidate	A	B	C	D	E	F	G	H
Judge 1 rank	3	1	7	5	2	8	4	6
Judge 2 rank	4	2	6	7	1	8	3	5

The Spearman rank correlation coefficient is $r_s = 0.929$. The critical value for a one-tailed test at the 5% level with $n = 8$ is 0.643.

(a) State suitable hypotheses for a one-tailed test of positive correlation. [1]

(b) Determine whether the result is significant, justifying your answer. [2]

(c) State one advantage of using Spearman's rather than Pearson's correlation coefficient for this data. [1]

Section C — Hard **HARD** (10 questions $\times$ 5 marks = 50 marks)

Scan me for help

21. **HARD** **Calculator** **[5 marks]**

A medical trial tests whether a new treatment increases the probability of recovery. Historically, 30% of patients recover without treatment. In the trial, 25 patients receive the new treatment.

Let X be the number of patients who recover. A one-tailed test is used with $H_0 : p = 0.3$, $H_1 : p > 0.3$, at the 5% significance level.

(a) Find the critical region for this test. [2]

(b) Calculate the exact probability of a Type I error for this test. [1]

(c) If the true probability of recovery with the new treatment is $p = 0.50$, find the probability of a Type II error. [2]

22. HARD Calculator **[5 marks]**

Vehicles pass through a rural checkpoint at a mean rate of 6 per hour. Following road works, the transport authority conducts a one-tailed test to detect whether the rate has fallen, observing the checkpoint for a single hour.

Let X be the number of vehicles observed. The test uses $H_0 : \lambda = 6$, $H_1 : \lambda < 6$ at the 5% significance level.

(a) Show that the critical region is $X \leq 1$. [2]

(b) Write down $P(\text{Type I error})$ for this test. [1]

(c) During road works, the true mean rate drops to $\lambda = 3.5$ vehicles per hour. Find $P(\text{Type II error})$. [2]

23. HARD Calculator **[5 marks]**

Two varieties of tomato plant (Variety A and Variety B) are compared. A horticulturist measures the heights (in cm) of plants in each group. Variety A: $n_A = 12$, $\bar{x}_A = 63.4$ cm, $s_A = 7.2$ cm. Variety B: $n_B = 15$, $\bar{x}_B = 57.8$ cm, $s_B = 5.9$ cm.

Assume both populations are normally distributed with equal variances.

(a) State the hypotheses for a one-tailed test that Variety A plants are taller on average. [1]

(b) Find the pooled standard deviation s_p, where $s_p^2 = \dfrac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2}$. [2]

(c) Calculate the p-value using your GDC and state the conclusion at the 5% significance level. [2]

24. HARD Calculator **[5 marks]**

The table gives the average monthly temperature (T, in $^{\circ}$C) and the monthly electricity consumption (E, in MWh) for a small hotel over a period of 9 months.

Month	1	2	3	4	5	6	7	8	9
T (°C)	5	7	10	14	18	23	27	25	20
E (MWh)	94	88	79	71	63	56	51	54	60

(a) Write down the value of the Pearson correlation coefficient r for this data. [1]

(b) Test at the 1% significance level whether there is a negative linear correlation between temperature and electricity consumption. State the hypotheses, p-value, and conclusion. [3]

(c) State one limitation of using a linear model to predict E from T for temperatures significantly above 30°C. [1]

25. HARD Calculator **[5 marks]**

An engineer tests whether the mean tensile strength of steel cables differs from a target value μ_0. A random sample of 16 cables gives $\bar{x} = 412$ MPa and $s = 18$ MPa.

(a) State the hypotheses for this two-tailed test. [1]

(b) The engineer reports a p-value of 0.0298. Using the formula $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$, find the value of μ_0 that produces this p-value. (You may use the fact that for 15 degrees of freedom the two-tailed p-value of 0.0298 corresponds to $|t| = 2.40$.) [2]

(c) Hence state the conclusion at the 5% significance level, in context. [1]

(d) Explain the effect on the p-value if the sample size were increased to $n = 64$ while $\bar{x}$, s, and μ_0 remain unchanged. [1]

26. **HARD** **Calculator** **[5 marks]**

A wildlife ecologist monitors a colony of 40 nesting birds. Under normal conditions, the probability that a randomly chosen egg hatches successfully is p. The ecologist tests $H_0 : p = p_0$ against $H_1 : p \neq p_0$ at the 5% significance level.

The critical region is found to be $\{X \leq 13\} \cup \{X \geq 26\}$, where X is the number of eggs that hatch.

(a) Write down the value of p_0. [1]

(b) Verify that $X \geq 26$ is the upper critical region by showing that $P(X \geq 26) \leq 0.025$ and $P(X \geq 25) > 0.025$ when $p_0 = 0.5$. [2]

(c) In a particular nesting season, 19 eggs hatch. State whether H_0 is rejected, giving a reason. [1]

(d) Find the probability of a Type I error for this test. [1]

27. HARD Calculator **[5 marks]**

A food technologist is testing whether the mean sodium content of a soup differs from the labelled value of $\mu_0 = 820$ mg per serving. A sample of $n = 20$ servings gives $\bar{x} = 836$ mg and $s = 29$ mg.

(a) Carry out a two-tailed t-test at the 5% significance level. State the hypotheses, p-value, and conclusion. [3]

(b) Explain what a Type I error would mean in this context. [1]

(c) The true mean sodium content is in fact 850 mg. State whether the result in (a) represents a Type II error. Justify your answer. [1]

28. **HARD** **Calculator** **[5 marks]**

A hospital records a mean of 2.4 patient falls per week on a particular ward. After introducing a new safety protocol, falls are monitored for a 4-week period. In those 4 weeks, 4 falls are recorded.

Let λ_0 be the mean number of falls in a 4-week period under the old protocol.

(a) Write down the value of λ_0. [1]

(b) State suitable hypotheses for testing whether the mean rate of falls has decreased. [1]

(c) Find the critical region for a one-tailed test at the 5% significance level, observing for a 4-week period. [1]

(d) Determine whether the result of 4 falls leads to rejection of H_0. State the conclusion in context. [1]

(e) Find the probability of a Type I error for this test. [1]

29. HARD Calculator **[5 marks]**

A geographer records the population density (people per km^2) and the mean house price (in $'000) for 10 districts. The raw data and ranks are given in the table below.

District	Density	Density rank	Price ($'000)	Price rank
1	120	1	145	1
2	280	2	162	2
3	510	3	178	3
4	740	4	199	5
5	890	5	193	4
6	1140	6	217	6
7	1560	7	238	7
8	2300	8	265	8
9	3800	9	289	9
10	9200	10	314	10

For this data: $r = 0.897$ (Pearson), $r_s = 0.988$ (Spearman).
Critical values for $n = 10$, one-tailed at 1%: Pearson 0.715, Spearman 0.745.

(a) Test at the 1% significance level whether population density and house price are positively correlated, using Pearson's correlation coefficient. State hypotheses, comparison, and conclusion. [2]

(b) Repeat the test using Spearman's rank correlation coefficient. [2]

(c) Explain which correlation coefficient is more appropriate for this data, referring to the nature of the data. [1]

30. HARD Calculator **[5 marks]**

A psychologist compares the recall scores (out of 20) of two groups of participants. Group P completed a mindfulness training programme; Group Q did not.

Group	Sample size	Mean score	Std dev
P (mindfulness)	22	14.3	3.1
Q (control)	20	12.6	3.5

(a) Carry out a two-tailed two-sample t-test at the 5% significance level to determine whether the mean recall scores differ between the two groups. State the hypotheses, p-value, and conclusion. (Do not assume equal variances.) [3]

(b) The psychologist also notes that 17 out of 22 participants in Group P score above the median of Group Q ($= 13$). Using a binomial model, test at the 5% significance level whether this proportion exceeds 0.5, and state the conclusion. [1]

(c) Explain how a Type I error in part (a) and a Type II error in part (b) would differ in their practical implications for the psychologist's conclusion about the programme's effectiveness. [1]

Worked Solutions — Section A (Easy)

Q1 — One-Sample t-test

(a) $H_0 : \mu = 250; \quad H_1 : \mu \neq 250$ **A1**

(where μ is the population mean energy content in kJ)

(b) Using GDC: $t = \dfrac{244 - 250}{12/\sqrt{18}} = -2.121\ldots; \quad p\text{-value} = 0.0487$ **(M1)A1**

Note: Award A1 for $p = 0.0487$ (accept 0.0486 to 0.0490).

(c) $p = 0.0487 < 0.05$, so we reject H_0. **R1**
There is sufficient evidence that the mean energy content differs from 250 kJ. **A1**

Note: Do not accept "the mean is not 250 kJ" without the word "evidence".

Q2 — Binomial Hypothesis Test

(a) Let p be the probability that a randomly chosen item is defective.
$H_0 : p = 0.2; \quad H_1 : p > 0.2$ **A1**

(b) $X \sim B(15, 0.2)$ under H_0.
p-value $= P(X \geq 6) = 1 - P(X \leq 5) = 1 - 0.9389\ldots = 0.0611\ldots \approx 0.0611$ **(M1)A1**

(c) $0.0611 < 0.10$, so we reject H_0. **R1**
There is sufficient evidence that the proportion of defective items exceeds 20%. **A1**

Q3 — Poisson Hypothesis Test

(a) Let λ be the mean number of complaints per hour.
$H_0 : \lambda = 4; \quad H_1 : \lambda > 4$ **A1**

(b) $X \sim$ Po(4) under H_0.
p-value $= P(X \geq 9) = 1 - P(X \leq 8) = 1 - 0.9786\ldots = 0.0214\ldots \approx 0.0214$ **(M1)A1**

(c) $0.0214 < 0.05$, so we reject H_0. **R1**
There is sufficient evidence that the mean complaint rate has increased. **A1**

Q4 — Correlation Test

(a) $H_0 : \rho = 0; \quad H_1 : \rho \neq 0$ **A1**

(b) $|r| = 0.743 > 0.576$ (critical value), so the result *is* significant. **R1A1**

(c) There is sufficient evidence of a (negative) linear correlation between altitude and mean daytime temperature. **A1**

Q5 — Type I Error

(a) A Type I error is rejecting H_0 when H_0 is in fact true. **A1**

(b) $P(\text{Type I error}) = 0.01$ (equal to the significance level). **A1**

(c) $0.008 < 0.01$, so we reject H_0. **A1**

Note: Full contextual conclusion not required here; accept "reject H_0".

Q6 — One-Sample t-test

(a) $H_0 : \mu = 380; \quad H_1 : \mu \neq 380$ **A1**

(b) $t = \dfrac{374 - 380}{14/\sqrt{10}} = -1.355\ldots; \quad p\text{-value} = 0.209\ldots \approx 0.209$ **(M1)A1**

(c) $0.209 > 0.10$, so we do not reject H_0. **R1**
There is insufficient evidence that the mean breaking strength differs from 380 N. **A1**

Q7 — Binomial Test (Lower Tail)

(a) Let p = probability a seed germinates.
$H_0 : p = 0.75; \quad H_1 : p < 0.75$ **A1**

(b) $X \sim B(20, 0.75)$ under H_0.
$p\text{-value} = P(X \leq 11) = 0.0409\ldots \approx 0.0409$ **(M1)A1**

(c) $0.0409 < 0.05$, so we reject H_0. **R1**
There is sufficient evidence that the germination rate is less than 75%. **A1**

Q8 — Type II Error

(a) A Type II error is failing to reject H_0 when H_0 is false (i.e. when the alternative hypothesis is true). **A1**

(b) The test does not reject $H_0 : \mu = 50$ but the true population mean is not 50 (i.e. $\mu \neq 50$). **A1**

(c) Increase the sample size n. (A larger sample reduces standard error, making the test more powerful and reducing the probability of a Type II error.) **A1**

Note: Accept any valid response, e.g. increase the significance level (with the caveat that this raises Type I error risk).

Q9 — Poisson Test (Lower Tail)

(a) Let λ = mean monthly accident rate.
$H_0 : \lambda = 3; \quad H_1 : \lambda < 3$ **A1**

(b) $X \sim \text{Po}(3)$ under H_0.
$p\text{-value} = P(X \leq 0) = P(X = 0) = e^{-3} = 0.0498\ldots \approx 0.0498$ **(M1)A1**

(c) $0.0498 < 0.05$, so we reject H_0. **R1**
There is sufficient evidence that the mean monthly accident rate has decreased. **A1**

Q10 — Pearson Correlation Test

(a) $H_0 : \rho = 0; \quad H_1 : \rho > 0$ **A1**

(b) $r = 0.997 > 0.748$ (critical value), so the result is significant. **R1A1**

(c) There is sufficient evidence of a positive linear correlation between hours studied and test score. **A1**

Worked Solutions — Section B (Medium)

Q11 — One-Sample t-test

(a) $H_0 : \mu = 1000; \quad H_1 : \mu \neq 1000$ (where μ = population mean volume, ml) **A1**

(b) The population of milk volumes must be normally distributed. **A1**

(c) $t = \dfrac{993 - 1000}{18/\sqrt{25}} = \dfrac{-7}{3.6} = -1.944\ldots$; p-value $= 0.0636\ldots \approx 0.0636$ **(M1)A1**

(d) $0.0636 > 0.05$, so we do not reject H_0. **R1**
There is insufficient evidence that the mean volume differs from 1000 ml. **A1**

Q12 — Binomial: Critical Region

(a) Let p = probability a frog is female.
$H_0 : p = 0.35; \quad H_1 : p \neq 0.35$ **A1**

(b) $X \sim B(20, 0.35)$ under H_0. Need $P(X \leq c_1) \leq 0.025$ and $P(X \geq c_2) \leq 0.025$: **M1**
Lower tail: $P(X \leq 2) = 0.0121 \leq 0.025; \quad P(X \leq 3) = 0.0444 > 0.025 \Rightarrow c_1 = 2$
Upper tail: $P(X \geq 12) = 0.0196 \leq 0.025; \quad P(X \geq 11) = 0.0480 > 0.025 \Rightarrow c_2 = 12$
Critical region: $X \leq 2$ or $X \geq 12$ **A1A1**

(c) Actual significance level $= P(X \leq 2) + P(X \geq 12)$
$= 0.01212\ldots + 0.01958\ldots = 0.03170\ldots \approx 3.17\%$ **A1**

Q13 — Poisson: Critical Region + Type I

(a) Let λ = mean number of meteors per night.
$H_0 : \lambda = 2.5; \quad H_1 : \lambda \neq 2.5$ **A1**

(b) $X \sim \text{Po}(2.5)$ under H_0. Need each tail ≤ 0.05: **M1**
Lower: $P(X = 0) = 0.0821 > 0.05$; so lower critical region is $\emptyset$ (no lower tail).
Upper: $P(X \geq 7) = 0.01416\ldots \leq 0.05; \quad P(X \geq 6) = 0.04202\ldots \leq 0.05; \quad P(X \geq 5) = 0.1088 > 0.05$
So upper critical value: $X \geq 6$
Critical region: $\{X \geq 6\}$ **A1A1**
Note: A two-tailed 10% test with a Poisson(2.5) distribution has no meaningful lower tail because $P(X = 0) = 0.0821 > 0.05$. The critical region is therefore one-sided in practice: $X \geq 6$.

(c) $P(\text{Type I error}) = P(X \geq 6 \mid \lambda = 2.5) = 0.0420\ldots \approx 0.0420$ **A1**

Q14 — Two-Sample t-test

(a) $H_0 : \mu_S = \mu_W; \quad H_1 : \mu_S \neq \mu_W$ (where μ_S, μ_W are population mean reaction times for sprinters and swimmers) **A1**

(b) Using GDC (2-SampTTest, pooled):
$t = -1.861\ldots;\quad p\text{-value} = 0.0713\ldots \approx 0.0713$ **(M1)A1**

(c) $0.0713 > 0.05$, so we do not reject H_0. **R1**
There is insufficient evidence that the mean reaction times differ between the two groups. **A1**

(d) Both populations are normally distributed (stated). **A1**
Accept: equal population variances.

Q15 — Pearson Correlation Test (Scatter Diagram)

(a) $H_0 : \rho = 0;\quad H_1 : \rho > 0$ **A1**

(b) $r = 0.9983 > 0.715$ (critical value), so the result is significant at the 1% level. **R1A1**
There is sufficient evidence of a positive linear correlation between advertising expenditure and monthly sales. **A1**

(c) The points on the scatter diagram lie very close to a straight line (i.e. the relationship appears to be linear). **R1**
Accept any equivalent observation, e.g. "the data shows a clear straight-line pattern".

Q16 — Binomial: Type I & Type II Errors

(a) $X \sim B(30, 0.9)$ under H_0.
$P(X \le 23) = 0.0480\ldots \le 0.05$ and $P(X \le 24) = 0.0988\ldots > 0.05$ **M1**
So the critical region is indeed $X \le 23$. **AG**

(b) $P(\text{Type I error}) = P(X \le 23 \mid p = 0.9) = 0.0480\ldots \approx 0.0480$ **A1**

(c) $X \sim B(30, 0.80)$ if true $p = 0.80$.
$P(\text{Type II error}) = P(X > 23 \mid p = 0.80) = P(X \ge 24 \mid p = 0.80)$ **M1**
$= 1 - P(X \le 23 \mid p = 0.80) = 1 - 0.5690\ldots = 0.431\ldots \approx 0.431$ **A1**

Q17 — One-Sample t-test: Find Sample Size

(a) $H_0 : \mu = 500;\quad H_1 : \mu \ne 500$ **A1**

(b) $p = 0.0310 < 0.05$, so we reject H_0. **R1**
There is sufficient evidence that the mean mass differs from 500 g. **A1**

(c) $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \dfrac{495 - 500}{8.4/\sqrt{n}} = -2.38$ **M1**

$\dfrac{-5\sqrt{n}}{8.4} = -2.38 \implies \sqrt{n} = \dfrac{2.38 \times 8.4}{5} = 3.9984 \approx 4 \implies n = 16$ **A1**

Note: Accept $n = 16$. With $n = 16$ and $t = -2.381$ (df $= 15$): $p \approx 0.0310$.

Q18 — Poisson Test (Two-Tailed, Rescaled)

(a) Let λ = mean number of faults per hour.
$H_0 : \lambda = 1.8; \quad H_1 : \lambda \neq 1.8$ **A1**

(b) Over a 2-hour period, the Poisson parameter is $\lambda_2 = 2 \times 1.8 = 3.6$. **A1**

(c) $X \sim$ Po(3.6) under H_0 for a 2-hour observation.
$P(X \geq 5) = 1 - P(X \leq 4) = 1 - 0.7064\ldots = 0.2936\ldots$
Since $H_1 : \lambda \neq 1.8$, the p-value for the upper tail: $2 \times P(X \geq 5) = 0.587\ldots$ **(M1)A1**
Note: $X = 5$ is above the mean of 3.6, so we compute the upper-tail p-value.
p-value ≈ 0.587 **A1**

(d) $0.587 > 0.05$, so we do not reject H_0. **R1**
There is insufficient evidence that the mean fault rate has changed after servicing. **A1**

Q19 — Two-Sample t-test (Unequal Variances)

(a) $H_0 : \mu_T = \mu_C; \quad H_1 : \mu_T \neq \mu_C$ (where μ_T, μ_C = population mean concentrations for tablet and capsule formulations) **A1**

(b) Using GDC (2-SampTTest, unequal variances / Welch's): $t = -1.826\ldots; \quad p$-value $= 0.0756\ldots \approx 0.0756$ **(M1)A1**

(c) $0.0756 > 0.05$, so we do not reject H_0. **R1**
There is insufficient evidence that the two formulations give different mean drug concentrations. **A1**

(d) The drug concentrations for each formulation are normally distributed in the population. **A1**

Q20 — Spearman Rank Correlation Test

(a) $H_0 : \rho_s = 0; \quad H_1 : \rho_s > 0$ **A1**

(b) $r_s = 0.929 > 0.643$ (critical value), so the result is significant at the 5% level. **R1A1**
There is sufficient evidence of a positive rank correlation between the two judges' rankings. **A1**

(c) Spearman's coefficient is appropriate because the data are *ordinal* (rankings), not continuous measurements. Pearson's requires continuous data with a linear relationship; Spearman's is valid for ranked/ordinal data. **R1**

Worked Solutions — Section C (Hard)

Q21 — Binomial: CR + Type I + Type II

(a) $X \sim B(25, 0.3)$ under H_0.
Need $P(X \geq c) \leq 0.05$: **M1**
$P(X \geq 12) = 1 - P(X \leq 11) = 1 - 0.9558\ldots = 0.0442\ldots \leq 0.05$
$P(X \geq 11) = 1 - P(X \leq 10) = 1 - 0.9022\ldots = 0.0978\ldots > 0.05$
Critical region: $X \geq 12$ **A1A1**

(b) $P(\text{Type I error}) = P(X \geq 12 \mid p = 0.3) = 0.0442\ldots \approx 0.0442$ **A1**

(c) $X \sim B(25, 0.5)$ if true $p = 0.5$.
$P(\text{Type II error}) = P(X < 12 \mid p = 0.5) = P(X \leq 11 \mid p = 0.5)$ **M1**
$= 0.3450\ldots \approx 0.345$ **A1**

Q22 — Poisson: CR + Type I + Type II

(a) $X \sim \text{Po}(6)$ under H_0. Need $P(X \leq c) \leq 0.05$: **M1**
$P(X \leq 1) = e^{-6}(1 + 6) = 7e^{-6} = 0.01735\ldots \leq 0.05$
$P(X \leq 2) = e^{-6}(1 + 6 + 18) = 25e^{-6} = 0.06197\ldots > 0.05$
So the largest c with $P(X \leq c) \leq 0.05$ is $c = 1$; hence the critical region is $X \leq 1$. **AG**

(b) $P(\text{Type I error}) = P(X \leq 1 \mid \lambda = 6) = 7e^{-6} = 0.01735\ldots \approx 0.0174$ **A1**

(c) $X \sim \text{Po}(3.5)$ if true $\lambda = 3.5$. CR is $X \leq 1$.
$P(\text{Type II error}) = P(X \notin \text{CR} \mid \lambda = 3.5) = P(X \geq 2 \mid \lambda = 3.5)$ **M1**
$= 1 - P(X \leq 1 \mid \lambda = 3.5) = 1 - e^{-3.5}(1 + 3.5) = 1 - 4.5e^{-3.5}$
$= 1 - 0.13589\ldots = 0.864\ldots \approx 0.864$ **A1**

Q23 — Two-Sample t-test with Pooled SD

(a) Let μ_A, μ_B = population mean heights of Variety A and Variety B.
$H_0 : \mu_A = \mu_B; \quad H_1 : \mu_A > \mu_B$ **A1**

(b) $s_p^2 = \dfrac{(12-1)(7.2)^2 + (15-1)(5.9)^2}{12 + 15 - 2} = \dfrac{11 \times 51.84 + 14 \times 34.81}{25}$ **M1**
$= \dfrac{570.24 + 487.34}{25} = \dfrac{1057.58}{25} = 42.303\ldots$
$s_p = \sqrt{42.303\ldots} = 6.504\ldots \approx 6.50$ cm **A1**

(c) Using GDC (2-SampTTest, pooled, one-tailed): p-value $= 0.0183\ldots \approx 0.0183$ **(M1)A1**
$0.0183 < 0.05$, so we reject H_0. **R1**
There is sufficient evidence that Variety A plants are taller on average than Variety B. **A1**

Q24 — Correlation Test + Regression + Limitation

(a) Using GDC on the data: $r = -0.9920\ldots \approx -0.992$ **A1**

(b) $H_0 : \rho = 0; \quad H_1 : \rho < 0$ **A1**
$|r| = 0.992$; critical value for one-tailed test at 1%, $n = 9$: $r_{\text{crit}} = 0.750$.
$0.992 > 0.750$, so the result is significant. **R1**
There is sufficient evidence of a negative linear correlation between temperature and electricity consumption. **A1**

(c) A linear model predicts that electricity consumption continues to decrease as temperature rises, but in reality, very high temperatures would increase consumption (e.g. due to air conditioning), so the linear model would be unreliable above $30°$C. **R1**

Q25 — One-Sample t-test: Recover μ_0

(a) $H_0 : \mu = \mu_0; \quad H_1 : \mu \neq \mu_0$ **A1**

(b) $t = \dfrac{412 - \mu_0}{18/\sqrt{16}} = \dfrac{412 - \mu_0}{4.5}$, and $|t| = 2.40$ **M1**

Since $\bar{x} = 412$ and the p-value $= 0.0298 < 0.05$ (test is significant), $\mu_0 \neq 412$.
$\dfrac{412 - \mu_0}{4.5} = \pm 2.40$
$\mu_0 = 412 \mp 2.40 \times 4.5 = 412 \mp 10.8$
$\mu_0 = 401.2$ or $\mu_0 = 422.8$
Both are valid (either root is possible); accept either value. **A1A1**

(c) $p = 0.0298 < 0.05$, so we reject H_0. **R1**
There is sufficient evidence that the mean tensile strength differs from the target value μ_0. **A1**

(d) With $n = 64$, $s/\sqrt{n} = 18/8 = 2.25$, so $|t| = \dfrac{|412 - \mu_0|}{2.25} = 4.80$. **M1**
The p-value decreases (the t-statistic becomes larger in magnitude for the same $\bar{x}$, s, μ_0), so the evidence against H_0 is stronger. **A1**

Q26 — Binomial: Recover p_0 from CR

(a) The critical region $\{X \leq 13\} \cup \{X \geq 26\}$ is symmetric about $np_0 = 20$, giving $p_0 = 20/40 = 0.5$. **A1**

(b) $X \sim B(40, 0.5)$ under H_0.
$P(X \geq 26) = 1 - P(X \leq 25)$ **M1**
$P(X \leq 25) = 0.9786\ldots$, so $P(X \geq 26) = 0.02136\ldots \leq 0.025$ ✓
$P(X \leq 24) = 0.9447\ldots$, so $P(X \geq 25) = 0.05530\ldots > 0.025$ ✓
Thus $X \geq 26$ is correct as the upper critical region. **A1A1**

(c) $X = 19$ does not lie in $\{X \leq 13\} \cup \{X \geq 26\}$, so H_0 is *not* rejected. There is insufficient evidence that $p \neq 0.5$. **A1**

(d) $P(\text{Type I error}) = P(X \leq 13) + P(X \geq 26)$

$= P(X \leq 13) + P(X \geq 26)$ [symmetric, so $= 2 \times P(X \geq 26)$]
$= 2 \times 0.02136\ldots = 0.04272\ldots \approx 0.0427$ **A1**

Q27 — One-Sample t-test + Type I/II Interpretation

(a) $H_0 : \mu = 820$; $\quad H_1 : \mu \neq 820$ $\quad$ (μ = population mean sodium content, mg) **A1**

$$t = \frac{836 - 820}{29/\sqrt{20}} = \frac{16}{6.485\ldots} = 2.467\ldots; \quad p\text{-value} = 0.0232\ldots \approx 0.0232$$ **(M1)A1**

$0.0232 < 0.05$, so we reject H_0. **R1**

There is sufficient evidence that the mean sodium content differs from 820 mg. **A1**

(b) A Type I error would mean concluding that the mean sodium content differs from 820 mg when it actually equals 820 mg. **A1**

(c) The result in (a) is to *reject* H_0. A Type II error is *failing* to reject H_0 when it is false. Since we rejected H_0, this cannot be a Type II error. (It could, however, be a Type I error.) **R1**

Q28 — Poisson: Rescale + CR + Type I

(a) $\lambda_0 = 4 \times 2.4 = 9.6$ **A1**

(b) Let λ = mean number of falls in a 4-week period.
$H_0 : \lambda = 9.6$; $\quad H_1 : \lambda < 9.6$ **A1**

(c) $X \sim \text{Po}(9.6)$ under H_0. Need $P(X \leq c) \leq 0.05$:
$P(X \leq 4) = 0.02701\ldots \leq 0.05$; $\quad P(X \leq 5) = 0.06548\ldots > 0.05$
Critical region: $X \leq 4$ **A1**

(d) $X = 4$ falls in the critical region $X \leq 4$, so we reject H_0. **R1**
There is sufficient evidence that the mean rate of falls has decreased after the safety protocol was introduced. **A1**

(e) $P(\text{Type I error}) = P(X \leq 4 \mid \lambda = 9.6) = 0.02701\ldots \approx 0.0270$ **A1**

Q29 — Pearson vs Spearman Correlation

(a) $H_0 : \rho = 0$; $\quad H_1 : \rho > 0$
$r = 0.897 > 0.715$ (critical value at 1%, $n = 10$), so the result is significant. **R1**
There is sufficient evidence of a positive linear correlation between population density and house price (Pearson). **A1**

(b) $H_0 : \rho_s = 0$; $\quad H_1 : \rho_s > 0$
$r_s = 0.988 > 0.745$ (critical value at 1%, $n = 10$), so the result is significant. **R1**
There is sufficient evidence of a positive rank correlation between population density and house price (Spearman). **A1**

(c) Spearman's coefficient is more appropriate because the population density data contains an extreme outlier (9200 vs the rest), making the data skewed and not approximately linear. Spearman's is resistant to outliers and valid for monotone (not necessarily linear) relationships. **R1**

Q30 — Two-Sample t + Binomial + Type I/II Synthesis

(a) $H_0 : \mu_P = \mu_Q$; $\quad H_1 : \mu_P \neq \mu_Q$ (where μ_P, μ_Q = population mean recall scores) **A1**
Using GDC (2-SampTTest, unequal variances):
$t = 1.695\ldots$; $\quad p$-value $= 0.0979\ldots \approx 0.0979$ **(M1)A1**
$0.0979 > 0.05$, so we do not reject H_0. **R1**
There is insufficient evidence of a difference in mean recall scores between the two groups. **A1**

(b) $Y \sim B(22, 0.5)$ under $H_0 : p = 0.5$.
$H_1 : p > 0.5$; $\quad p$-value $= P(Y \geq 17 \mid p = 0.5) = 1 - P(Y \leq 16) = 1 - 0.9267\ldots = 0.0733\ldots > 0.05$ **(M1)**
$0.0733 > 0.05$, so we do not reject H_0. There is insufficient evidence that the proportion exceeds 0.5. **A1**

(c) A Type I error in (a) would mean incorrectly concluding that the programme makes a difference when it does not — the psychologist would over-state the programme's effectiveness, potentially recommending an ineffective intervention.
A Type II error in (b) would mean failing to detect that a higher proportion of Group P outperformed the Group Q median, even when a real effect exists — the psychologist would under-state (miss) evidence of the programme's benefit. **R1**

Common Mistakes to Avoid

1. **Wrong tail direction.** A claim that a parameter has *increased* requires $H_1 : \theta > \theta_0$ (upper tail). Using a two-tailed test when the question says "greater than" or "less than" loses the conclusion mark.
2. **Conclusion wording.** Never write "we accept H_0" or "we prove H_1". Use "there is sufficient/insufficient evidence that …".
3. **Poisson rescaling.** If the test observes a period of length t but the stated rate is per unit time λ_0, the parameter under H_0 becomes $\lambda_0 t$. Forgetting to rescale is a very common error.
4. **Discrete critical regions.** Because X is integer-valued, the exact tail probability may not equal α. Always check *both* k and $k \pm 1$ to identify the correct boundary.
5. **Type II error direction.** A Type II error is *failing to reject* H_0 when H_1 is true. It is not the same as the p-value. Compute $P(X \notin \text{CR} \mid \text{true parameter})$.
6. **Pearson vs Spearman.** Pearson's r requires roughly linear data without severe outliers. For ranked/ordinal data or data with outliers, use Spearman's r_s. Check both the scatter plot and the context.
7. **Two-sample t: equal variances.** Unless the question states "assume equal variances", use Welch's t-test (Pooled: No on GDC). Using pooled when not instructed loses marks.