

MATH40005

Probability and Statistics

Revised Mock Examination 2 - Solutions and Marking Guide

Time allowed	2 hours
Questions	4
Total marks	80
Format	Closed book; justify all answers
Difficulty calibration	A 32, B 20, C 12, D 16
Paper role	2025 historical bridge: every high-mark 2025 ability family receives direct or method-level training, recalibrated to the 2026 2-hour paper.

- Answer all four questions.
- Start each question on a new page in an actual timed attempt.
- State assumptions and named results used.
- Statistical tables or stated critical values may be used where supplied.

Question 1

1(a) (5 marks)

Recognition signal	Route categories form a partition and route-specific rates are conditional probabilities.
First key step	Write $P(L)=\sum_r P(L r)P(r)$.
Historical basis	2025 Q1(b)(i)
Ability/method	law of total probability with observational weights; K03

Course-method solution. By the law of total probability, $P(L)=0.04(0.5)+0.10(0.3)+0.05(0.2)=0.06$.

Marking points. 1 mark: notation/partition; 2 marks: LTP; 2 marks: calculation

Common errors. Averaging the late rates without route weights.

Final self-check. The weighted result 0.06 lies between 0.04 and 0.10.

Difficulty allocation: A:4, B:1; estimated normal time 6 min.

1(b) (3 marks)

Recognition signal	The conditioning direction is reversed from the provided late rate.
First key step	Use the total probability from part (a) as the denominator.
Historical basis	2025 Q1(b)(ii)
Ability/method	Bayes reversal after total probability; K08

Course-method solution. Bayes gives $P(A|L)=P(L|A)P(A)/P(L)=0.04(0.5)/0.06=1/3$.

Marking points. 1 mark: Bayes formula; 1 mark: numerator; 1 mark: result

Common errors. Reporting $P(L|A)=0.04$.

Final self-check. $P(A|L)=1/3$ is less than the prior $P(A)=1/2$ because A has the lowest late rate.

Difficulty allocation: A:2, B:1; estimated normal time 4 min.

1(c) (4 marks)

Recognition signal	An intervention changes conditional probabilities, so the denominator must be rebuilt.
First key step	Update the data-generating mechanism before reusing Bayes.
Historical basis	2025 Q1(b)(iii)
Ability/method	Bayes under a changed mechanism; K08,K03

Course-method solution. Now $P(L)=0.04(0.5)+1(0.3)+1(0.2)=0.52$. Therefore $P(A|L)=0.02/0.52=1/26$.

Marking points. 1 mark: update conditional rates; 1 mark: new total probability; 1 mark: Bayes; 1 mark: result

Common errors. Using the old $P(L)=0.06$.

Final self-check. The posterior should fall sharply because almost all B/C trains are late.

Difficulty allocation: B:1, C:1, D:2; estimated normal time 6 min.

1(d) (5 marks)

Recognition signal	Same combinatorial architecture as a multitype sequence.
First key step	Choose positions, then values.
Historical basis	2025 Q2(a)(i)
Ability/method	multitype password count; K05

Course-method solution. The count is $C(10,2)C(8,5)(5^*4)52^5 10^3$.

Marking points. 2 marks positions; 1 mark distinct-symbol ordering; 2 marks product

Common errors. Missing the order of symbols.

Final self-check. Check against $10!/(5!3!2!) C(5,2)2!52^5 10^3$.

Difficulty allocation: A:2, B:2, C:1; estimated normal time 6 min.

1(e) (3 marks)

Recognition signal	Plot selection depends on whether the goal is a relationship, robust distribution summary, or time order.
First key step	Classify the variable types and the analytical purpose.
Historical basis	2023 Q6(d)(i); 2023 Q6(d)(iii); 2023 Q6(d)(iv); 2026 Q3(c)(i)-(ii)
Ability/method	plot choice for relationship, robust summary and time series; K47

Course-method solution. (i) A scatterplot. (ii) A boxplot. (iii) A time-series/line plot in chronological order.

Marking points. 1 mark each plot with purpose

Common errors. Using a histogram for a two-variable relationship; ignoring chronology in a stock series.

Final self-check. Each chosen plot directly displays the feature named in the question.

Difficulty allocation: A:2, B:1; estimated normal time 4 min.

Total: 20 marks

Question 2

2(a) (4 marks)

Recognition signal	A monotone transform of a uniform variable suggests the CDF method.
First key step	Solve $-\log U/\lambda \leq t$ carefully, noting the inequality direction.
Historical basis	2025 Q3(a)(i)
Ability/method	inverse-transform/CDF method; K20,K24

Course-method solution. For $t < 0$, $F_T(t) = 0$. For $t \geq 0$, $F_T(t) = P(U \geq e^{-\lambda t}) = 1 - e^{-\lambda t}$. Thus $T \sim \text{Exponential}(\lambda)$.

Marking points. 1 mark support; 2 marks CDF transformation; 1 mark identification

Common errors. Writing $P(U \leq e^{-\lambda t})$.

Final self-check. The density is $\lambda e^{-\lambda t}$ on $t > 0$.

Difficulty allocation: A:3, B:1; estimated normal time 5 min.

2(b) (4 marks)

Recognition signal	A maximum event becomes an intersection.
First key step	Write $\{M \leq m\} = \{X_1 \leq m, \dots, X_n \leq m\}$.
Historical basis	2025 Q3(a)(ii)
Ability/method	order-statistic CDF; K21

Course-method solution. For $0 < m < 1$, $F_M(m) = P(\text{all } X_i \leq m) = m^n$. It is 0 for $m \leq 0$ and 1 for $m \geq 1$. The density is $n m^{n-1}$ on $(0,1)$.

Marking points. 1 mark event rewrite; 1 mark independence; 1 mark piecewise CDF; 1 mark density

Common errors. Using $1 - (1-m)^n$, which is for a minimum-related event.

Final self-check. $F_M(1) = 1$ and the density integrates to 1.

Difficulty allocation: A:3, B:1; estimated normal time 5 min.

2(c) (4 marks)

Recognition signal	The exponential tilt of a normal density is handled by completing the square.
First key step	Combine exponents before integrating.
Historical basis	2025 Q3(b)(i)
Ability/method	MGF derivation by completing the square; K12,K24

Course-method solution. $M_X(t) = \int \exp(tx) (2\pi)^{-1/2} \exp(-x^2/2) dx$. Complete the square: $tx - x^2/2 = -(x-t)^2/2 + t^2/2$. The shifted normal density integrates to 1, so $M_X(t) = \exp(t^2/2)$, t real.

Marking points. 1 mark setup; 2 marks completing square; 1 mark result/domain

Common errors. Treating the integral as elementary without normalization.

Final self-check. $M_X(0) = 1$ and M_X is even.

Difficulty allocation: A:1, B:1, C:1, D:1; estimated normal time 7 min.

2(d) (4 marks)

Recognition signal	A sum of independent variables suggests multiplying MGFs.
First key step	Use the result from part (c).
Historical basis	2025 Q3(b)(ii)
Ability/method	MGF product and distribution identification; K12,K24

Course-method solution. Independence gives $M_S(t) = \prod \exp(t^2/2) = \exp(nt^2/2)$, the MGF of $N(0,n)$. Therefore $S \sim N(0,n)$.

Marking points. 1 mark independence/product; 1 mark MGF; 1 mark identification; 1 mark parameters

Common errors. Giving variance $\sqrt{n}$.

Final self-check. The mean is 0 and variance is n .

Difficulty allocation: A:1, B:1, C:1, D:1; estimated normal time 4 min.

2(e) (4 marks)

Recognition signal	Moment equality alone need not determine a distribution.
First key step	Add a condition ensuring the moment series defines the MGF near zero.
Historical basis	2025 Q3(c)(i)
Ability/method	moment determinacy via MGF existence; K34,K12

Course-method solution. A sufficient condition is that both MGFs exist on an open interval containing 0. Then their Taylor series have coefficients $E[U^k]/k!$ and $E[V^k]/k!$, so the MGFs agree near 0. Uniqueness of MGFs implies U and V have the same distribution.

Marking points. 1 mark condition; 2 marks equality of MGFs; 1 mark uniqueness conclusion

Common errors. Claiming moments always determine the distribution.

Final self-check. The argument must invoke a uniqueness theorem.

Difficulty allocation: B:1, C:1, D:2; estimated normal time 6 min.

Total: 20 marks

Question 3

3(a) (2 marks)

Recognition signal	Different means do not affect variance.
First key step	Apply variance to the fixed linear combination.
Historical basis	2025 Q4(a)(i)
Ability/method	variance of an independent sample mean; K26,K30

Course-method solution. By independence, $\text{Var}(\bar{X}) = n^{-2} \sum \text{Var}(X_i) = n^{-2} \sum \sigma^2 = \sigma^2/n$.

Marking points. 1 mark variance scaling; 1 mark independence sum

Common errors. Adding standard deviations.

Final self-check. For $n=1$ the formula reduces to σ^2 .

Difficulty allocation: A:2; estimated normal time 3 min.

3(b) (6 marks)

Recognition signal	The usual unbiasedness fails when the means differ.
First key step	Rewrite the centered sum before taking expectations.
Historical basis	2025 Q4(a)(ii)
Ability/method	expectation of sample variance under unequal means; K30,K31

Course-method solution. Use $\sum (X_i - \bar{X})^2 = \sum X_i^2 - n \bar{X}^2$. Now $E[\sum X_i^2] = \sum (\sigma^2 + \mu_i^2) = n \sigma^2 + \sum \mu_i^2$. Also $E[\bar{X}^2] = \sigma^2/n + \bar{\mu}^2$. Hence $E[(n-1)S^2] = (n-1)\sigma^2 + \sum (\mu_i - \bar{\mu})^2$, so $E[S^2] = \sigma^2 + [\sum (\mu_i - \bar{\mu})^2]/(n-1)$.

Marking points. 1 mark decomposition; 2 marks expectations; 2 marks algebra; 1 mark final expression

Common errors. Assuming $E[S^2] = \sigma^2$; omitting between-mean variation.

Final self-check. If all μ_i are equal, the extra term vanishes.

Difficulty allocation: A:1, B:2, C:1, D:2; estimated normal time 9 min.

3(c) (4 marks)

Recognition signal	Unknown variance and normality imply a t interval.
First key step	Use s , not s^2 .
Historical basis	2025 Q4(b)(i)
Ability/method	Student-t confidence interval; K36

Course-method solution. The interval is $\bar{x} \pm 1.796 s/\sqrt{12}$. Since $s = \sqrt{3}$, the margin is 0.898, so (180.102, 181.898).

Marking points. 1 mark pivot/df; 1 mark standard error; 1 mark margin; 1 mark interval

Common errors. Using 1.645 normal critical value.

Final self-check. Endpoints are symmetric around 181.

Difficulty allocation: A:3, B:1; estimated normal time 5 min.

3(d) (4 marks)

Recognition signal	Variance inference uses a chi-square pivot and reverses endpoints when inverted.
First key step	Write the probability statement before solving for σ^2 .
Historical basis	2025 Q4(b)(ii)
Ability/method	chi-square variance interval; K37

Course-method solution. Since $11S^2/\sigma^2 \sim \chi^2_{11}$, the interval is $(11 \cdot 3/21.920, 11 \cdot 3/3.816) = (1.505, 8.648)$.

Marking points. 1 mark pivot; 1 mark inversion/order; 1 mark substitution; 1 mark result

Common errors. Putting the larger chi-square quantile in the upper denominator.

Final self-check. The lower bound is smaller than s^2 and the upper bound larger.

Difficulty allocation: A:2, B:1, C:1; estimated normal time 6 min.

3(e) (4 marks)

Recognition signal	A Q-Q plot is judged by systematic departures from the reference line.
First key step	Inspect center and tails separately.
Historical basis	2025 Q4(d); 2023 Q6(d)(ii)
Ability/method	Q-Q diagnostic; K47

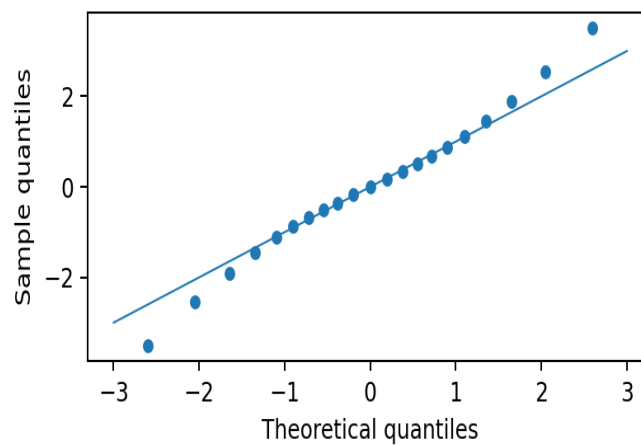
Course-method solution. The middle points are near the line but both tails deviate outward. This is evidence of heavier tails than a normal distribution, so the plot does not support a satisfactory normal model.

Marking points. 1 mark pattern; 1 mark tail interpretation; 1 mark conclusion; 1 mark qualified wording

Common errors. Requiring every point to be exactly on the line; ignoring tail curvature.

Final self-check. Heavy tails produce more extreme sample quantiles than the normal line.

Difficulty allocation: A:1, C:1, D:2; estimated normal time 4 min.



Total: 20 marks

Question 4

4(a) (5 marks)

Recognition signal	Both endpoints enter the support and the likelihood decreases with interval width.
First key step	Write constraints first, then shrink the interval to the sample range.
Historical basis	2025 Q6(a)
Ability/method	two-boundary support-dependent MLE; K54

Course-method solution. The likelihood is $(\theta_2 - \theta_1)^{-n}$ when $\theta_1 \leq x_{\min}$ and $\theta_2 \geq x_{\max}$, and zero otherwise. To maximize it, minimize the interval length subject to containing all data: $\theta_1_{\hat{}} = x_{\min}$ and $\theta_2_{\hat{}} = x_{\max}$.

Marking points. 1 mark likelihood/support; 2 marks optimization argument; 2 marks estimators

Common errors. Differentiating inside an unconstrained region; choosing sample mean.

Final self-check. The fitted support must contain every observation exactly with minimum width.

Difficulty allocation: A:1, B:1, C:1, D:2; estimated normal time 7 min.

4(b) (4 marks)

Recognition signal	Writing estimators as linear combinations makes covariance algebra direct.
First key step	Use $\sum(x_i - \bar{x}) = 0$.
Historical basis	2025 Q6(b)(i)-(iii)
Ability/method	regression-estimator covariance; K49, K48

Course-method solution. Because $\sum(x_i - \bar{x}) = 0$, $S_{xy} = \sum(x_i - \bar{x})(y_i - \bar{y}) = \sum(x_i - \bar{x})y_i$. Thus $\beta_{1\hat{}} = \sum a_i y_i$ with $a_i = (x_i - \bar{x})/S_{xx}$ and $\beta_{0\hat{}} = \bar{y} - \bar{x} \beta_{1\hat{}}$. $\text{Cov}(\bar{y}, \beta_{1\hat{}}) = 0$ and $\text{Var}(\beta_{1\hat{}}) = \sigma^2/S_{xx}$, so $\text{Cov}(\beta_{0\hat{}}, \beta_{1\hat{}}) = -\bar{x} \sigma^2/S_{xx}$.

Marking points. 1 mark identity; 1 mark linear-estimator form; 1 mark covariance pieces; 1 mark result

Common errors. Assuming the estimators are independent; dropping $\bar{x}$.

Final self-check. If $\bar{x} = 0$, the covariance is 0.

Difficulty allocation: A:1, B:1, C:1, D:1; estimated normal time 7 min.

4(c) (4 marks)

Recognition signal	Bootstrap approximates the sampling distribution by resampling the empirical distribution.
First key step	Resample observations, not residuals, for this simple statistic.
Historical basis	2025 Q6(c)
Ability/method	percentile bootstrap interval; K52

Course-method solution. Draw many independent resamples of size n with replacement from the observed values. Compute the median for each resample. Use the empirical 2.5th and 97.5th percentiles of the bootstrap medians as the percentile interval.

Marking points. 1 mark resample size/replacement; 1 mark replicate statistic; 1 mark repetition; 1 mark quantiles

Common errors. Sampling without replacement; using original-data quantiles rather than median replicates.

Final self-check. The interval endpoints are quantiles of the bootstrap statistic distribution.

Difficulty allocation: A:1, B:2, C:1; estimated normal time 5 min.

4(d) (7 marks)

Recognition signal	Known and unequal population variances signal a two-sample z test, not a pooled t test.
First key step	Write $\text{Var}(\bar{X} - \bar{Y}) = \sigma_1^2/n + \sigma_2^2/m$.
Historical basis	2025 Q5(b)
Ability/method	known-variance two-sample z test with p-value; K40

Course-method solution. Under H_0 , $Z = (\bar{X} - \bar{Y}) / \sqrt{10/15 + 4/12} = -3 / \sqrt{1} = -3$. The two-sided p-value is $2\Phi(-3) = 0.00270$. Since $0.00270 < 0.01$, reject H_0 ; the data provide evidence that the population mean heights differ. The assumptions are independent normal samples and the stated known variances.

Marking points. 1 mark hypotheses/assumptions; 2 marks standard error/statistic; 2 marks p-value; 1 mark comparison; 1 mark qualified conclusion

Common errors. Pooling the known variances; using a one-sided p-value; saying the means are proved unequal.

Final self-check. The standard error is exactly 1 and the observed z-score is -3.

Difficulty allocation: A:2, B:1, C:1, D:3; estimated normal time 10 min.

Total: 20 marks