

MATH40005

Probability and Statistics

Revised Mock Examination 6 - 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	Seven-year mixed transfer pressure paper: combines distinct historical families, changes information order and context, and preserves the 2026 total difficulty.

- 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) (6 marks)

Recognition signal	The factor $n-k$ signals a distinguished member among a complement set.
First key step	Identify the binomial coefficient identity $nC(n-1,k)=(n-k)C(n,k)$.
Historical basis	2022 Q1(b) method, altered parameter
Ability/method	algebraic and narrative double counting; K05

Course-method solution. Algebra: $n(4+1)^{n-1} = \sum nC(n-1,k)4^k = \sum C(n,k)(n-k)4^k$. Story: n people choose among four non-water drinks and water, with one designated driver who drinks water. Count by choosing driver first, or by choosing k non-water drinkers and then the driver among $n-k$ water drinkers.

Marking points. 3 marks algebra; 3 marks common-object story

Common errors. Giving only algebra or only a story.

Final self-check. Check $n=1$.

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

1(b) (6 marks)

Recognition signal	Comparing posteriors under changed prevalence tests understanding of prior/base-rate effects.
First key step	Write Bayes separately for each prevalence.
Historical basis	2022 Q3(a); 2024 Q1(c)
Ability/method	Bayes/base-rate transfer; K08

Course-method solution. At prevalence 0.01: $0.95(0.01)/[0.95(0.01)+0.05(0.99)]=0.1610$. At 0.10: $0.095/[0.095+0.045]=0.6786$. The likelihood characteristics are unchanged, but the prior odds increase; Bayes updates prior odds by the same likelihood ratio.

Marking points. 2 marks first posterior; 2 marks second; 2 marks explanation

Common errors. Holding the denominator fixed.

Final self-check. Posterior odds equal prior odds times 19.

Difficulty allocation: A:2, B:2, C:1, D:1; estimated normal time 8 min.

1(c) (5 marks)

Recognition signal	Exact one-pair counting requires excluding further collisions.
First key step	Choose the pair first.
Historical basis	2023 Q1(b)
Ability/method	exact birthday collision; K07

Course-method solution. The probability is $C(12,2)*365*(364)_{-10}/365^{12}$.

Marking points. 1 mark pair; 1 mark common day; 2 marks distinct remaining days; 1 mark denominator

Common errors. Using the complement of no collision, which includes multiple collisions.

Final self-check. The numerator has 12 labeled people throughout.

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

1(d) (3 marks)

Recognition signal	Independent finite stages form a Cartesian product.
First key step	Use ordered pairs.
Historical basis	2025 Q1(a)(i); 2021 Q1(b)
Ability/method	finite product probability space; K01

Course-method solution. $\Omega=\{1,2,3,4\} \times \{H,T\}$, $F=P(\Omega)$, and each singleton has probability $1/8$, so $P(A)=|A|/8$.

Marking points. 1 mark each component

Common errors. Omitting the event space.

Final self-check. There are 8 equally likely outcomes.

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

Total: 20 marks

Question 2

2(a) (8 marks)

Recognition signal	A conditional expectation from a joint density needs normalization, marginalization and a conditional integral.
First key step	Compute f_X before forming the conditional density.
Historical basis	2022 Q3(c)(i)-(iii)
Ability/method	conditional expectation on triangular support; K17,K18,K23

Course-method solution. Normalization gives $c=24/7$. For $0 < x < 1$, $f_X(x) = c x [3/2 - x - x^2/2]$. The conditional density is $(y+1)/[3/2 - x - x^2/2]$ for $x < y < 1$. Hence $E[Y|X=x] = [5/6 - x^3/3 - x^2/2]/[3/2 - x - x^2/2]$, $0 < x < 1$.

Marking points. 2 marks c ; 2 marks marginal; 1 mark conditional density; 2 marks integral; 1 mark domain

Common errors. Using y from 0 to 1 rather than x to 1.

Final self-check. The conditional density integrates to 1.

Difficulty allocation: A:4, B:2, C:1, D:1; estimated normal time 11 min.

2(b) (7 marks)

Recognition signal	Random sum $\rightarrow$ condition on N .
First key step	Use both total expectation and total variance.
Historical basis	2024 Q3(b)(ii)-(iii)
Ability/method	compound-Poisson moments; K13,K23

Course-method solution. $E[S|N] = Nm$ and $\text{Var}(S|N) = Nv$. Therefore $E[S] = \lambda m$ and $\text{Var}(S) = E[Nv] + \text{Var}(Nm) = \lambda(v + m^2)$.

Marking points. 2 marks conditional moments; 2 marks mean; 3 marks total variance

Common errors. Omitting $\text{Var}(E[S|N])$.

Final self-check. For deterministic $Y_i = 1$, recover $\text{Poi}(\lambda)$.

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

2(c) (5 marks)

Recognition signal	Combine an order-statistic CDF with a decreasing transformation.
First key step	Handle the inequality reversal under $-\log$.
Historical basis	2025 Q3(a); historical transformation chain
Ability/method	order statistic followed by monotone transform; K21,K20,K24

Course-method solution. For $t \geq 0$, $F_T(t) = P(M \geq e^{-t}) = 1 - P(M < e^{-t}) = 1 - e^{-nt}$. Thus $T \sim \text{Exponential}(n)$. For $t < 0$ the CDF is 0.

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

Common errors. Writing $e^{-t/n}$ as the survival rather than 1 minus it.

Final self-check. Mean should be $1/n$: larger n makes M closer to 1 and T smaller.

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

Total: 20 marks

Question 3

3(a) (6 marks)

Recognition signal	Unequal means create extra between-mean variation.
First key step	Use the sum-of-squares identity.
Historical basis	2025 Q4(a)(i)-(ii)
Ability/method	non-identical-normal sample variance expectation; K30

Course-method solution. Using $\sum (X_i - \bar{X})^2 = \sum X_i^2 - n\bar{X}^2$ gives $E[S^2] = \sigma^2 + \sum (\mu_i - \bar{\mu})^2 / (n-1)$.

Marking points. 2 marks decomposition; 2 marks expectations; 2 marks result

Common errors. Applying the i.i.d. unbiasedness theorem.

Final self-check. Equal means remove the extra term.

Difficulty allocation: A:1, B:2, C:1, D:2; estimated normal time 8 min.

3(b) (7 marks)

Recognition signal	The tested linear combination is a sum, but independent sample means still add variances.
First key step	Standardize $\bar{X} + \bar{Y} - 360$ using the pooled variance.
Historical basis	2023 Q5(b)
Ability/method	pooled two-sample t test for a linear combination; K40

Course-method solution. The pooled variance is 4. Under H_0 , $T = (\bar{X} + \bar{Y} - 360) / [s_p \sqrt{1/n + 1/m}] = (357.5 - 360) / [2 \sqrt{1/16 + 1/9}] = -3.00$ with 23 df. The two-sided p-value is about 0.0064, so reject H_0 .

Marking points. 1 mark hypotheses/assumptions; 1 mark pooled variance; 2 marks statistic/df; 2 marks p-value; 1 mark conclusion

Common errors. Using $\bar{X} - \bar{Y}$; using df $n + m$.

Final self-check. The absolute statistic 3 is clearly beyond a typical 5% two-sided threshold.

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

3(c) (4 marks)

Recognition signal	A high-confidence variance interval is wide and asymmetric.
First key step	Invert the chi-square inequalities carefully.
Historical basis	2026 Q3(d)(i)
Ability/method	chi-square variance interval; K37

Course-method solution. The interval is $(11 \cdot 3 / 26.757, 11 \cdot 3 / 2.603) = (1.233, 12.678)$.

Marking points. 1 mark pivot; 1 mark correct quantiles/order; 1 mark substitution; 1 mark result

Common errors. Swapping endpoints.

Final self-check. The interval contains $s^2 = 3$.

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

3(d) (3 marks)

Recognition signal	Continuity permits the probability integral transform without randomisation.
First key step	Write p as a function of the null CDF.
Historical basis	2024 Q5(a)(i); 2026 Q3(e)
Ability/method	p-value definition and probability-integral-transform proof; K41

Course-method solution. For observed t , the upper-tail p-value is $p = 1 - F_0(t)$, the null probability of a result at least as large. Under H_0 , $F_0(T) \sim \text{Uniform}(0,1)$ by the probability integral transform; therefore $1 - F_0(T)$ is also $\text{Uniform}(0,1)$. Thus $P_{H_0}(p \leq \alpha) = \alpha$.

Marking points. 1 mark definition; 1 mark probability integral transform; 1 mark uniform conclusion

Common errors. Calling p the probability H_0 is true; stating uniformity without proof.

Final self-check. The result implies $P(p \leq \alpha) = \alpha$ under H_0 .

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

Total: 20 marks

Question 4

4(a) (5 marks)

Recognition signal	A one-parameter linear-in-beta model is minimized by a normal equation.
First key step	Differentiate with respect to beta.
Historical basis	2024 Q6(c)
Ability/method	least squares for a transformed predictor; K48

Course-method solution. $RSS(\beta) = \sum (y_i - \beta x_i)^2$. Differentiate: $-2 \sum x_i^2 (y_i - \beta x_i) = 0$, giving $\hat{\beta} = \frac{\sum x_i^2 y_i}{\sum x_i^4}$ (provided the denominator is positive). The second derivative is $2 \sum x_i^4 > 0$.

Marking points. 2 marks RSS/derivative; 2 marks estimator; 1 mark minimum condition

Common errors. Using $\sum x_i y_i / \sum x_i^2$.

Final self-check. Units and the second derivative check the formula.

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

4(b) (3 marks)

Recognition signal	Residuals should have approximately constant spread around zero.
First key step	Compare the vertical spread at small and large x.
Historical basis	2020 Q6(b) residual-diagnostic method; 2026 Q4(b)
Ability/method	heteroscedastic residual diagnosis and variance-stabilising remedy; K50

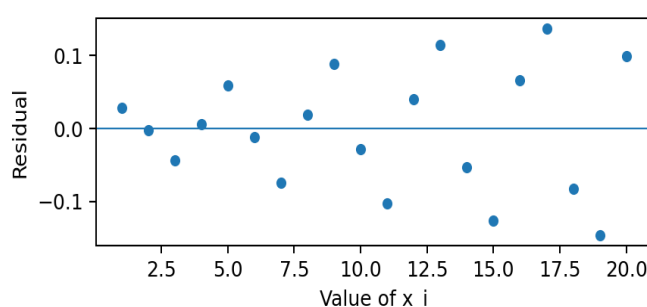
Course-method solution. The residual spread increases with x, giving a fan shape. This indicates non-constant error variance (heteroscedasticity), so the constant-variance regression assumption is doubtful. A variance-stabilising response transformation or an appropriately weighted regression may improve the fit; residuals must then be checked again.

Marking points. 1 mark fan/increasing spread; 1 mark heteroscedasticity conclusion; 1 mark justified remedy

Common errors. Diagnosing nonlinear mean curvature when the central trend remains near zero; relying only on R^2 .

Final self-check. After a successful remedy, the residual spread should be roughly constant across x.

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



4(c) (5 marks)

Recognition signal	The likelihood is flat on the feasible support interval.
First key step	Translate all data constraints simultaneously.
Historical basis	2026 Q4(f)
Ability/method	non-unique support MLE; K54

Course-method solution. The likelihood is 1 when $\max x_{i-1} \leq \theta \leq \min x_i$ and zero otherwise. Hence all θ in $[x_{\max-1}, x_{\min}]$ are MLEs, provided the observed range is at most 1.

Marking points. 1 mark likelihood; 2 marks support inequalities; 1 mark MLE set; 1 mark feasibility

Common errors. Choosing only $x_{\min}$.

Final self-check. The interval length equals $1-(x_{\max}-x_{\min})$.

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

4(d) (4 marks)

Recognition signal	Geometric likelihood is conjugate to a beta prior.
First key step	Collect exponents of θ and $1-\theta$.
Historical basis	2021 Q6(b)
Ability/method	Beta-geometric conjugacy; K55

Course-method solution. The likelihood is $\theta^n(1-\theta)^{\sum x_i}$. Multiplying by $\theta^{a-1}(1-\theta)^{b-1}$ gives $\text{Beta}(a+n, b+\sum x_i)$.

Marking points. 1 mark likelihood; 1 mark kernel; 1 mark parameters; 1 mark identification

Common errors. Using $b+\sum x_i$ rather than $b+\sum x_i - n$.

Final self-check. Both posterior parameters remain positive for valid data.

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

4(e) (3 marks)

Recognition signal	Time trends can create spurious correlation.
First key step	Separate descriptive association from causal inference.
Historical basis	2024 Q5(d); 2026 Q4(d)
Ability/method	spurious correlation interpretation; K51

Course-method solution. The sample shows strong association, but neither causation nor a substantive relationship follows. Shared time trends, confounding or coincidence can produce a spurious correlation; causal evidence requires more than r .

Marking points. 1 mark association; 1 mark no causation; 1 mark explanation

Common errors. Saying correlation proves no relation; it does describe the sample.

Final self-check. Detrending or design information would be needed for stronger claims.

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

Total: 20 marks