End of Topic Test – Chapter 5: The Normal Distribution — Mark Scheme

Total: 39 marks

1. Understanding the normal distribution

Heights $\sim N(165, 7^2)$.

(a) B1 symmetric bell-shaped curve drawn, centred and labelled at mean 165 cm. B1 points at one standard deviation each side correctly labelled, 158 cm and 172 cm. *[2]*

(b) B1 recognise 158 and 172 are exactly ±1 standard deviation from the mean. B1 proportion = 68% (the empirical/68–95–99.7 rule). *[2]*

(c) B2 (or B1+B1) valid explanation, e.g. adult height results from many small independent genetic and environmental factors combining, which tends to produce a symmetric, bell-shaped distribution clustered around a mean – a good match to the normal model. *[2]*

2. Calculating probabilities

$X \sim N(50, 12^2)$.

(a) M1 $z = \dfrac{70-50}{12} = 1.667$. M1 use normal tables/calculator for $P(Z > 1.667)$. A1 $P(X > 70) \approx 0.0478$ (accept 0.047–0.048). *[3]*

(b) M1 both z-values: $z_1 = \dfrac{40-50}{12} = -0.833$, $z_2 = \dfrac{65-50}{12} = 1.25$. M1 $P(40 < X < 65) = \Phi(1.25) - \Phi(-0.833)$. A1 ≈ 0.692 (69.2%). *[3]*

(c) B2 (or B1+B1) correct contextual interpretation, e.g. approximately 69% (about 7 in 10) of students taking this test are expected to score between 40 and 65. *[2]*

3. Running times

Time $\sim N(27.4, 4.1^2)$.

(a) M1 $z = \dfrac{25-27.4}{4.1} = -0.585$. M1 use tables/calculator. A1 $P(X < 25) \approx 0.279$ (27.9%). *[3]*

(b) M1 recognise the fastest 10% corresponds to the lower tail, using $z = -1.2816$ (invNorm(0.10)). M1 time $= 27.4 + (-1.2816)(4.1)$. A1 threshold ≈ 22.1 minutes (accept 22.0–22.2). *[3]*

(c) B2 (or B1+B1) valid explanation, e.g. the normal distribution is unbounded and symmetric, but real race times cannot be negative and there is likely a realistic practical upper limit (e.g. a course time-out/cut-off, or runners who walk), so the true distribution of the slowest runners is likely skewed rather than symmetric like the normal model. *[2]*

4. Interpreting a normal distribution diagram

Lifetimes $\sim N(1200, 150^2)$.

(a) B1 recognise 1050 and 1350 are ±1 standard deviation from the mean. B1 proportion = 68%. *[2]*

(b) M1 $z = \dfrac{1500-1200}{150} = 2.0$. M1 use tables/calculator for $P(Z > 2)$. A1 ≈ 0.0228 (2.28%). *[3]*

(c) B2 (or B1+B1) valid reason, e.g. the normal model is simple to apply and gives a good approximation over most of the range (especially near the mean), so remains useful for practical decisions (e.g. setting warranty periods) even if the extreme tails are not perfectly modelled. *[2]*

5. Interpreting a study using the normal distribution

IQ $\sim N(100, 15^2)$.

(a) B2 (or B1+B1) "normally distributed" means the data forms a symmetric, bell-shaped distribution about the mean (100), with spread described by the standard deviation (15); values close to the mean are most common, and the further a value is from the mean, the less common it is. *[2]*

(b) M1 $z = \dfrac{130 - 100}{15} = 2.0$. M1 use tables/calculator for $P(Z > 2)$. A1 ≈ 0.0228 (2.28%). *[3]*

(c) B1 the quotation notes real data show deviations from the idealised model due to sampling variation, measurement error and differing testing conditions. B1 this means the model is only an approximation, and this is most relevant for extreme (very high/low) scores. B1 conclusion: probabilities calculated from the idealised normal model (e.g. part (b)) may not be perfectly accurate for real individuals, especially at the extremes. *[3]*

(d) B2 (or B1+B1) valid reason, e.g. fewer people achieve very high/low scores, so factors like test anxiety, inconsistent test administration or limited sample diversity have a proportionally larger effect on those extreme scores, making them less reliable than scores near the mean. *[2]*