Chapter 5: Normal Distribution — Mark Scheme

Throughout, $z = \dfrac{x-\mu}{\sigma}$ and $\Phi(z) = P(Z < z)$ for the standard normal distribution $N(0,1)$, with $\Phi(-z) = 1 - \Phi(z)$. Standard normal table values are quoted to 4–5 d.p. and z-values to 2 d.p., consistent with a typical AQA-supplied normal distribution table.

5.1 Features of a normal distribution

Activity 1. $X \sim N(80, 5^2)$ (rod length, mm); tube holds up to 90 mm.

a) "At least 10 mm spare room" means length ≤ 80 mm, i.e. $z = \dfrac{80-80}{5} = 0$. $P(X \leq 80) = \Phi(0) = 0.5$. Expected number $= 100\,000 \times 0.5 = \mathbf{50\,000}$ rods.

b) Too long means $X > 90$: $z = \dfrac{90-80}{5} = 2$. $P(X > 90) = 1 - \Phi(2) = 1 - 0.9772 = 0.0228$. Expected number $= 100\,000 \times 0.0228 \approx \mathbf{2280}$ rods.

Exercise 5A

1. $X \sim N(65, 7^2)$ (weight, kg). The table gives three weights (72, 51, 79) for which the number of s.d. above the mean must be found, and two s.d. values (-0.5, 1.5) for which the corresponding weight must be found.

- 72 kg: $z = \dfrac{72-65}{7} = \mathbf{1}$
- 51 kg: $z = \dfrac{51-65}{7} = \mathbf{-2}$
- 79 kg: $z = \dfrac{79-65}{7} = \mathbf{2}$
- $z = -0.5$: weight $= 65 + (-0.5)(7) = \mathbf{61.5}$ kg
- $z = 1.5$: weight $= 65 + (1.5)(7) = \mathbf{75.5}$ kg

2.

a) Roughly symmetric, single-peaked and bell-shaped $\Rightarrow$ **this does appear normal**.

b) Heavily skewed with a peak near 0 and a long tail $\Rightarrow$ **not normal** – a normal distribution must be symmetric about its mean, but this distribution is skewed (not symmetric) with unequal tails.

3. Using the fact that about 95% of the data ("the bulk") lies within 2 s.d. of the mean, half of the stated range $\approx 2\sigma$.

a) Mean $= 12$; half-range $= (15-9)/2 = 3 = 2\sigma \Rightarrow \sigma = \mathbf{1.5}$.

b) Mean $= 50$; half-range $= (62-38)/2 = 12 = 2\sigma \Rightarrow \sigma = \mathbf{6}$.

c) Mean $= 100$; half-range $= (160-40)/2 = 60 = 2\sigma \Rightarrow \sigma = \mathbf{30}$.

d) Mean $= 1200$; half-range $= (1260-1140)/2 = 60 = 2\sigma \Rightarrow \sigma = \mathbf{30}$.

4. Female: $N(0.25, 0.03^2)$; Male: $N(0.22, 0.025^2)$.

a) Women have a slower average reaction time (mean 0.25 s vs 0.22 s) and their reaction times are more spread out/variable (s.d. 0.03 s vs 0.025 s).

b) Female: $z = \dfrac{0.31-0.25}{0.03} = \dfrac{0.06}{0.03} = 2$. Male: $z = \dfrac{0.27-0.22}{0.025} = \dfrac{0.05}{0.025} = 2$. Both values standardise to the same $z = 2$, so (since probabilities for the standard normal distribution depend only on z) the two proportions are identical: $P(Z > 2) = 1 - \Phi(2) = 1 - 0.9772 = \mathbf{0.0228}$ (2.28%).

5. Daily returns $\sim N(0.04\%, 0.20\%^2)$.

a) The mean of 0.04% shows that, historically, the fund has made a very small positive average daily return (slight net growth per day on average).

b) The s.d. (0.20%) is five times the size of the mean, so day-to-day returns are highly volatile relative to the average gain – on many days the actual return will be well away from (and could easily be negative compared to) the mean.

c) i) $z = \dfrac{-0.36 - 0.04}{0.20} = \dfrac{-0.40}{0.20} = -2$. $P(Z < -2) = 1 - \Phi(2) = 0.0228$. Expected days per year $\approx 0.0228 \times 365 \approx \mathbf{8}$ days.

ii) $z = \dfrac{0.64 - 0.04}{0.20} = \dfrac{0.60}{0.20} = 3$. $P(Z > 3) = 1 - \Phi(3) = 1 - 0.9987 = 0.0013$. Expected days per year $\approx 0.0013 \times 365 \approx \mathbf{0.5}$, i.e. less than 1 day (about 0–1 days) a year.

6.

a) So that customer orders can be met immediately/without delay, since demand varies day to day and production/delivery is not instant.

b) Holding excess stock ties up capital, incurs storage costs, and risks the stock becoming obsolete, damaged or unsellable.

c) Sales $\sim N(500, 25^2)$. $z = \dfrac{450 - 500}{25} = -2$. $P(X < 450) = \Phi(-2) = 1 - \Phi(2) = 1 - 0.9772 = \mathbf{0.0228}$ (2.28%), so it is unlikely.

7. Pasta weight $\sim N(500, 4^2)$, $n = 300$.

a) $z = \dfrac{495 - 500}{4} = -1.25$, $z = \dfrac{505 - 500}{4} = 1.25$. $P(495 < X < 505) = 2\Phi(1.25) - 1 = 2(0.8944) - 1 = 0.7887$. $P(\text{outside}) = 1 - 0.7887 = 0.2113$. Expected number $= 300 \times 0.2113 \approx \mathbf{63}$ packets.

b) $z = \dfrac{497.5 - 500}{4} = -0.625$, $z = \dfrac{502.5 - 500}{4} = 0.625$. $P = 2\Phi(0.625) - 1 = 2(0.7340) - 1 = 0.4680$. Expected number $= 300 \times 0.4680 \approx \mathbf{140}$ packets.

c) $z = \dfrac{497.5 - 500}{4} = -0.625$ and 500 is the mean ($z = 0$). $P(497.5 < X < 500) = \Phi(0) - \Phi(-0.625) = 0.5 - (1 - 0.7340) = 0.2340$. Expected number $= 300 \times 0.2340 \approx \mathbf{70}$ packets.

8. Calls/minute $\sim N(25, 6^2)$.

a) $z = \dfrac{40 - 25}{6} = 2.5$. $P(X > 40) = 1 - \Phi(2.5) = 1 - 0.9938 = \mathbf{0.0062}$ (0.62%).

b) $z = \dfrac{15 - 25}{6} = -1.67$. $P(X < 15) = 1 - \Phi(1.67) = 1 - 0.9525 = \mathbf{0.0475}$ (4.75%).

c) $z = \pm 1.67$. $P(15 < X < 35) = 2\Phi(1.67) - 1 = 2(0.9525) - 1 = \mathbf{0.9050}$ (90.5%).

5.2 Calculating probabilities

Activity 2. By symmetry of the standard normal curve about $z = 0$: $\Phi(0) = \mathbf{0.5}$; $\Phi(1) \approx \mathbf{0.8413}$; $\Phi(2) \approx \mathbf{0.9772}$; $\Phi(-1) = 1 - \Phi(1) \approx \mathbf{0.1587}$; $\Phi(-2) = 1 - \Phi(2) \approx \mathbf{0.0228}$.

Activity 3.

a) $P(z = 1) = \mathbf{0}$ – for a continuous distribution the probability of any single exact value is zero (probability is area, and a single point has zero width).

b) $P(-0.4 \leq z \leq 1.7) = \Phi(1.7) - \Phi(-0.4) = 0.9554 - (1 - 0.6554) = 0.9554 - 0.3446 = \mathbf{0.6109}$.

c) Probability of an impossible event $= \mathbf{0}$.

d) Probability of a certain event $= \mathbf{1}$.

Exercise 5B

1.

a) i) $P(-1 < z < 1) = 2\Phi(1) - 1 = 2(0.8413) - 1 = \mathbf{0.6827}$

ii) $P(-2 < z < 2) = 2\Phi(2) - 1 = 2(0.9772) - 1 = \mathbf{0.9545}$

iii) $P(-3 < z < 3) = 2\Phi(3) - 1 = 2(0.9987) - 1 = \mathbf{0.9973}$

b) These match the empirical rule quoted in the reference box: $0.6827 \approx 68\%$ (close to the stated $\frac{2}{3}/67\%$), $0.9545 \approx 95\%$, and $0.9973 \approx 1$ ("virtually all"), confirming the 68–95–99.7 rule.

2.

a) $P(z > 1.4) = 1 - \Phi(1.4) = 1 - 0.9192 = \mathbf{0.0808}$

b) $P(z < -1.6) = 1 - \Phi(1.6) = 1 - 0.9452 = \mathbf{0.0548}$

c) $P(0.8 < z < 2.2) = \Phi(2.2) - \Phi(0.8) = 0.9861 - 0.7881 = \mathbf{0.1980}$

5.3 Standardising a normal variable

Activity 4. Test scores $\sim N(500, 110^2)$; claim is that scoring > 700 equals the top 4%. $z = \dfrac{700-500}{110} = \dfrac{200}{110} = 1.82$. $P(Z > 1.82) = 1 - \Phi(1.82) = 1 - 0.9656 = 0.0344$ (3.44%). Since $3.44\% \neq 4\%$, scoring over 700 is **not exactly** the same as being in the top 4% – it is actually a slightly higher bar, corresponding to roughly the top 3.4% of test-takers, so the website's claim is only approximately true.

Exercise 5C

1. σ is the square root of the variance (second parameter of $N(\mu, \sigma^2)$).

a) $N(8, 6^2)$: $\sigma = \mathbf{6}$

b) $N(8, 49)$: $\sigma = \sqrt{49} = \mathbf{7}$

c) $N(21, 60)$: $\sigma = \sqrt{60} \approx \mathbf{7.75}$

2. $z = \dfrac{x-\mu}{\sigma}$.

a) $z = \dfrac{11-7}{4} = \mathbf{1}$

b) $z = \dfrac{5-9}{12} = \mathbf{-0.33}$

c) $z = \dfrac{-2-3}{5} = \mathbf{-1}$

d) $z = \dfrac{7-5}{4} = \mathbf{0.5}$ (since $N(5,16) \Rightarrow \sigma = 4$)

e) $z = \dfrac{2-4}{3} = \mathbf{-0.67}$ (since $N(4,9) \Rightarrow \sigma = 3$)

f) $z = \dfrac{6-8}{\sqrt{8}} = \mathbf{-0.71}$ (since $N(8,8) \Rightarrow \sigma = \sqrt{8} \approx 2.83$)

3.

a) $P(z < 1.32) = \Phi(1.32) = \mathbf{0.9066}$

b) $P(z > 1.87) = 1 - \Phi(1.87) = 1 - 0.9693 = \mathbf{0.0307}$

c) $P(z < -2) = 1 - \Phi(2) = 1 - 0.9772 = \mathbf{0.0228}$

d) $P(z > -4) = \Phi(4) \approx \mathbf{1.0000}$ (virtually certain)

e) $P(0.5 < z < 2.1) = \Phi(2.1) - \Phi(0.5) = 0.9821 - 0.6915 = \mathbf{0.2907}$

f) $P(-2 < z < 1) = \Phi(1) - \Phi(-2) = 0.8413 - 0.0228 = \mathbf{0.8186}$

g) $P(-2.18 < z < -1.33) = \Phi(2.18) - \Phi(1.33) = 0.9854 - 0.9082 = \mathbf{0.0771}$

4.

a) By symmetry of the standard normal curve about $z = 0$, the area to the left of $-a$ is the mirror image (in the line $z = 0$) of the area to the right of a. Reflecting the sketch about $z = 0$ maps the tail $z < -a$ exactly onto the tail $z > a$, so the two areas (probabilities) are equal: $P(z < -a) = P(z > a)$.

b) $P(-a < z < a) = P(z < a) - P(z < -a) = P(z < a) - P(z > a)$ (using part a). Since $P(z > a) = 1 - P(z < a)$, this gives $P(-a < z < a) = P(z < a) - (1 - P(z < a)) = \mathbf{2P(z < a) - 1}$.

5. All values of a found from $\Phi(a) =$ given probability (interpolating in the table).

a) $\Phi(a) = 0.722 \Rightarrow a \approx \mathbf{0.58}$

b) $\Phi(a) = 1 - 0.02 = 0.98 \Rightarrow a \approx \mathbf{2.05}$

c) $\Phi(a) = 0.15 < 0.5 \Rightarrow a < 0$. Since $\Phi(1.04) \approx 0.85 = 1 - 0.15$, $a \approx -\mathbf{1.04}$

d) $2\Phi(a) - 1 = 0.90 \Rightarrow \Phi(a) = 0.95 \Rightarrow a \approx \mathbf{1.64}$ (more precisely 1.645)

e) $2\Phi(a) - 1 = 0.50 \Rightarrow \Phi(a) = 0.75 \Rightarrow a \approx \mathbf{0.67}$

6. Typing speed $R \sim N(72, 12)$, so $\sigma = \sqrt{12} \approx 3.464$.

a) $2\Phi(a) - 1 = 0.34 \Rightarrow \Phi(a) = 0.67 \Rightarrow a \approx 0.44$. $c = a\sigma = 0.44 \times 3.464 \approx 1.52$. So $72 - c \approx \mathbf{70.48}$ and $72 + c \approx \mathbf{73.52}$.

b) $2\Phi(a) - 1 = 0.754 \Rightarrow \Phi(a) = 0.877 \Rightarrow a \approx 1.16$. $c = 1.16 \times 3.464 \approx 4.02$. So $72 - c \approx \mathbf{67.98}$ and $72 + c \approx \mathbf{76.02}$.

7.

a) A sharp spike with little spread can occur when the value is constrained by an external fixed limit (e.g. a legal/regulatory limit, a rounding or measurement resolution, or a strong physical/biological constraint), so nearly all individuals cluster tightly around it rather than varying continuously.

b) Two separate peaks (bimodality) typically arise when the population is really a mixture of two distinct sub-groups, each with its own different typical value (e.g. combining two different categories, such as males and females, or two different processes/conditions), producing two separate clusters instead of one.

Consolidation Exercise 5

1. Kitten weight $\sim N(1.2, 0.18^2)$; large $\Leftrightarrow$ weight > 1.5 kg; $P(\text{small}) = 0.137$.

a) $z = \dfrac{1.5 - 1.2}{0.18} = \dfrac{0.3}{0.18} = 1.67$. $P(\text{large}) = P(X > 1.5) = 1 - \Phi(1.67) = 1 - 0.9525 = \mathbf{0.0475}$ (4.75%).

b) $P(X < k) = 0.137 \Rightarrow \Phi\left(\dfrac{k-1.2}{0.18}\right) = 0.137$. Since $0.137 < 0.5$, the required z is negative; $\Phi(1.09) \approx 0.863 = 1 - 0.137$, so $z \approx -1.09$. $k = 1.2 + (-1.09)(0.18) = 1.2 - 0.196 = \mathbf{1.00}$ kg (maximum weight classed as small).

2. Dog temperature $\sim N(38.5, 0.4^2)$. $z = \dfrac{40-38.5}{0.4} = \dfrac{1.5}{0.4} = 3.75$. $P(X > 40) = 1 - \Phi(3.75) \approx 1 - 0.9999 = \mathbf{0.0001}$ (about 0.01%, i.e. very unlikely).

3. Girls' heights $\sim N(133.5, 11.2^2)$.

a) $z = \dfrac{136-133.5}{11.2} = 0.22$. $P(X > 136) = 1 - \Phi(0.22) = 1 - 0.5871 = \mathbf{0.4129}$

b) $z = \dfrac{138-133.5}{11.2} = 0.40$. $P(X < 138) = \Phi(0.40) = \mathbf{0.6554}$

c) $z = \dfrac{128-133.5}{11.2} = -0.49$. $P(X < 128) = \Phi(-0.49) = 1 - 0.6879 = \mathbf{0.3121}$

d) $z = \dfrac{132-133.5}{11.2} = -0.13$, $z = \dfrac{135-133.5}{11.2} = 0.13$. $P(132 < X < 135) = 2\Phi(0.13) - 1 = 2(0.5517) - 1 = \mathbf{0.1034}$

4. Idris's time $X \sim N(18, 6^2)$.

a) $z = \dfrac{25-18}{6} = 1.17$. $P(X < 25) = \Phi(1.17) = \mathbf{0.8790}$

b) $z = \dfrac{15-18}{6} = -0.5$. $P(X > 15) = 1 - \Phi(-0.5) = \Phi(0.5) = \mathbf{0.6915}$

5. Hana's install time $X \sim N(35, 15^2)$.

a) $z = \dfrac{45-35}{15} = 0.67$. $P(X < 45) = \Phi(0.67) = \mathbf{0.7486}$

b) $z = \dfrac{20-35}{15} = -1$, $z = \dfrac{45-35}{15} = 0.67$. $P(20 < X < 45) = \Phi(0.67) - \Phi(-1) = 0.7486 - 0.1587 = \mathbf{0.5899}$

c) $\Phi(z) = 0.9 \Rightarrow z \approx 1.28$. $k = 35 + 1.28 \times 15 = 35 + 19.2 = \mathbf{54.2}$ minutes.

6. Ankle circumference $\sim N(225, 18^2)$.

a) $z = \dfrac{250-225}{18} = 1.39$. $P(X \leq 250) = \Phi(1.39) = 0.9177 \approx \mathbf{91.8}\%$

b) $z = \dfrac{195-225}{18} = -1.67$, $z = \dfrac{245-225}{18} = 1.11$. $P(195 < X < 245) = \Phi(1.11) - \Phi(-1.67) = 0.8665 - 0.0475 = 0.8190 \approx \mathbf{81.9}\%$

7. Blood pressure $\sim N(122, 15^2)$.

a) i) $z = \dfrac{130-122}{15} = 0.53$. $P(X > 130) = 1 - \Phi(0.53) = 1 - 0.7019 = \mathbf{0.2981}$

ii) $z = \dfrac{110-122}{15} = -0.8$, $z = \dfrac{125-122}{15} = 0.2$. $P(110 < X < 125) = \Phi(0.2) - \Phi(-0.8) = 0.5793 - 0.2119 = \mathbf{0.3674}$

b) $\Phi(z) = 0.9 \Rightarrow z \approx 1.28$. BP $= 122 + 1.28 \times 15 = 122 + 19.2 = \mathbf{141.2}$