Maths Solutions (Statistics)

Section A: Multiple Choice

1. **Answer:** 0.7

   Working:
   Median: 4.9
   Lower quartile: 4.7
   Upper quartile: 5.4
   IQR = 5.4 - 4.7 = 0.7

2. **Answer:** 25

   Working:
   Sampling interval = population size / sample size
   = 1000 / 40
   = 25

3. **Answer:** 7/39

   Working:
   6 women + 7 men = 13 runners
   Total number of teams of 8 is ${}^{13}C_8 = 1287$
   Teams with more women - possible combinations are:
   6 women + 2 men : ${}^{6}C_6 * {}^{7}C_2 = 1 * 21 = 21$, or
   5 women + 3 men : ${}^{6}C_5 * {}^{7}C_3 = 6 * 35 = 210$
   Probability = (21 + 210) / 1287 = 231/1287 = 7/39

4. **Answer:** Each trial must have equally likely outcomes.

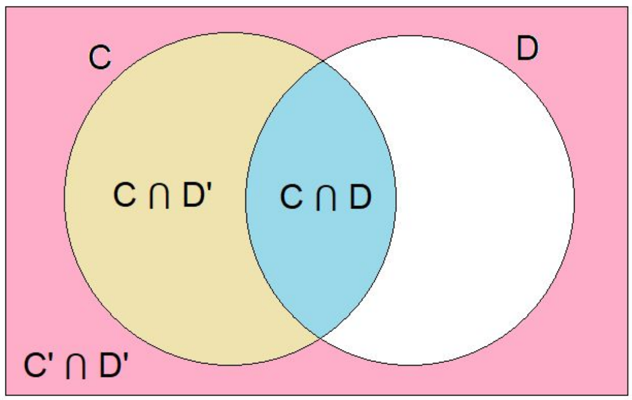
   Working: The binomial model requires two independent mutually exclusive outcomes with constant probabilities.

5. **Answer:** Mean = 2.8, Variance = 1.68

Working: Mean of binomial variable = np = 7 * 0.4 = 2.8
Variance of binomial variable = np(1 - p) = 7 * 0.4 * 0.6 = 1.68

6. **Answer:** $(C' \cap D)'$

Working: Draw a Venn diagram and shade all the different sections:



The remaining region is C’ ∩ D, so the event region is (C’ ∩ D)’.

7. **Answer:** {0, 1, 2, 3} ∪ {15, 16, 17, …, 24, 25}

Working: Critical region = rejection region → probability < 2.5% in each tail (two tail hypothesis test)
Left tail: P(X ≤ 3) = 0.00968 and P(X ≤ 4) = 0.032 → {0, 1, 2, 3}
Right tail: P(X ≥ 14) = 0.0255 and P(X ≥ 15) = 0.00931
→ {15, 16, 17, …, 25}

8. **Answer:** 50 years

Working: Sum of ages originally = 28 * 20 = 560
Sum of ages after = 30 * 22 = 660
Age added = 660 - 560 = 100
Mean age added = 100 / 2 = 50

9a. **Answer:** The 2065 households in the village

Working: The population is the set of all possible individuals that can be sampled from.

9b. **Answer:** Systematic

Working: Quota: setting sample strata (sub-group) sizes beforehand
Systematic: sampling at specified intervals
Stratified: sampling in proportions from different strata
Simple random: using random number generator

Section B: Standard Questions

10.

a. Total area = (20 * 0.3) + (20 * 1.3) + (10 * 2) + (10 * 1.6) + (40 * 0.8) = 100 [1 mark]
→ 250 / 100 = 2.5 students per unit area [1 mark]
Area of interest = (7 * 2) + (12 * 0.8) = 23.6 [1 mark]
→ 23.6 * 2.5 = 59 students [1 mark] (Accept answers 57-61 students inclusive)

b. Find the area where the left-hand area = 50 (half of 100):
(20 * 0.3) + (20 * 1.3) = 32 [1 mark]
Next area has 20, so need 18 more to get to 50 → 9/10 of it
9/10 of class 40-50 is at 49
→ median is (approximately) 49 [1 mark] (Accept answers 48-50 inclusive but **not** 50 if no valid work shown)

c. Constructing a table,

midpoint	frequency (= f.d. * c.w. * 2.5)
10	20 * 0.3 * 2.5 = 15
30	20 * 1.3 * 2.5 = 65
45	10 * 2 * 2.5 = 50
55	10 * 1.6 * 2.5 = 40
80	40 * 0.8 * 2.5 = 80
	total = 250

Using calculator in stats mode, $\sum x = 12950$, $\sum x^2 = 794250$, $n = 250$

i) $\bar{x} = \sum x / n = 12950 / 250 = 51.8$ [2 marks]

ii) $\sigma = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$ $= \sqrt{(794250 / 250) - (51.8)^2} = 22.2$ [2 marks]

11.

a. $\sum x_n = 140 + (40 * 50) = 2140$ [1 mark]

$\sum (x_n - 50)^2 = \sum (x_n^2 - 100x_n + 2500) = 4490$ [1 mark]
$\rightarrow \sum x_n^2 = 4490 - (2500 * 40) + 100 \sum x_n$ [1 mark]
$\rightarrow \sum x_n^2 = 4490 - (2500 * 40) + (100 * 2140) = 118490$ [1 mark]

$\mu = \sum x_n / n = 2140 / 40 = 53.5$ [1 mark]

$\sigma = \sqrt{\frac{\sum x^2}{n} - \mu^2}$ $= \sqrt{((118490 / 40) - (53.5)^2)} = 10$ [1 mark]

b. 0 [1 mark]

(Since $\sum (x_n - \bar{x}) = \sum x_n - n\bar{x} = n\bar{x} - n\bar{x} = 0$)

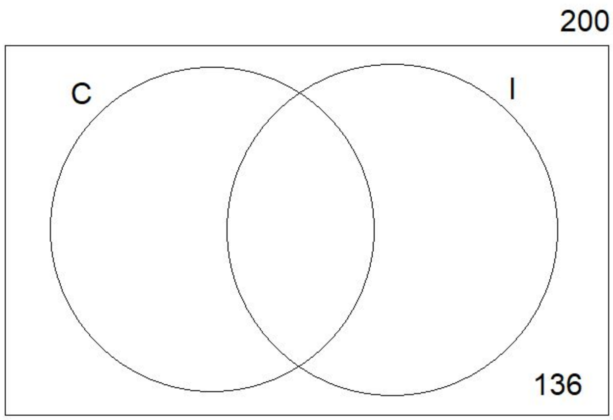
12. For the whole class:
$\bar{x} = 65 \rightarrow \sum x = 65 * 40 = 2600$ [1 mark]
$\sigma = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$ $= 18 \rightarrow \sqrt{((\sum x^2) / 40 - 65^2)} = 18 \rightarrow \sum x^2 = 181960$ [2 marks]

For the boys:
$\bar{x}_B = 72 \rightarrow \sum x = 72 * 24 = 1728$ [1 mark]
$\sigma = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$ $= 20 \rightarrow \sqrt{((\sum x_B^2) / 24 - 72^2)} = 20 \rightarrow \sum x_B^2 = 134016$ [2 marks]

Sum of boys + Sum of girls = Sum of class
$\rightarrow \sum x_G = 2600 - 1728 = 872$
$\rightarrow \sum x_G^2 = 181960 - 134016 = 47944$ [1 mark]

$\rightarrow \bar{x}_G = 872 / 16 = 54.5$ [1 mark]
$\rightarrow$ $\sigma = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$ $= \sqrt{((47944 / 16) - 54.5^2)} = 5.12$ [1 mark]

13. Venn diagram:

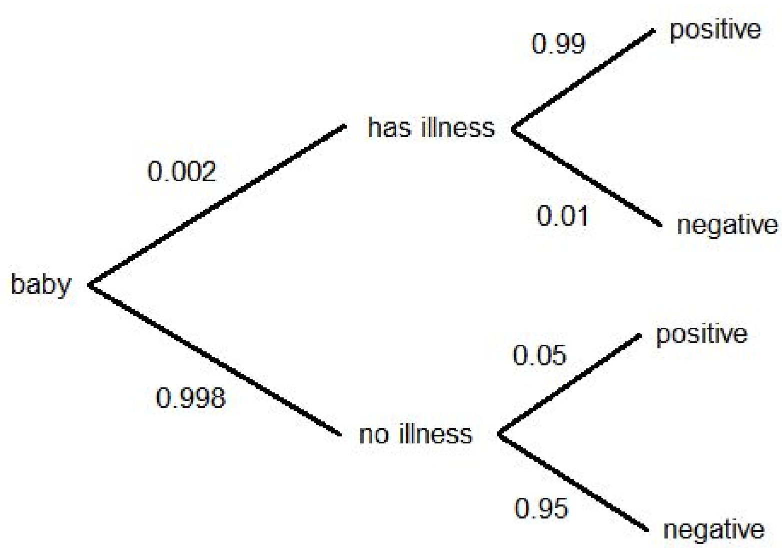


$P(C \mid I) = P(C \cap I) / P(I) \rightarrow P(C \cap I) = P(C \mid I) * P(I)$
$\rightarrow P(C \cap I) = 0.4 * P(I)$ [1 mark]

$P((C \cup I)') = 136/200 \rightarrow P(C \cup I) = 64/200 = 8/25$ [1 mark]
$P(C \cup I) = P(C) + P(I) - P(I \cap C)$ [1 mark]
$8/25 = 0.23 + P(I) - 0.4 * P(I)$ [1 mark]
$0.09 = 0.6 * P(I)$ [1 mark]
$\rightarrow P(I) = 0.15$ [1 mark]
$\rightarrow P(C \cap I) = 0.4 * 0.15 = 0.06$ [1 mark]

14.

a. [1 mark for first branch, 1 mark for second branches]



b. i) (0.002 * 0.99) + (0.998 * 0.05) [1 mark] = 0.05188 = 0.0519 [1 mark]

ii) (0.002 * 0.99) + (0.998 * 0.95) [1 mark] = 0.95008 = 0.9501 [1 mark]

c. P(ill | positive) = P(ill ∩ positive) / P(positive) [1 mark]
P(ill ∩ positive) = 0.002 * 0.99 = 0.00198 [1 mark]
P(positive) = 0.0519
→ P(ill | positive) = 0.00198 / 0.0519 = 0.03815 = 0.038 (2 s.f.) [1 mark]

The probability that a positive result is actually ill is very low, so the test is not very useful. [1 mark]

d. P(ill | positive) = P(ill ∩ positive) / P(positive) = 0.5 [1 mark]
P(ill ∩ positive) = 0.002 * 0.99 = 0.00198 [1 mark]
P(positive) = 0.00198 + (0.998 * (1 - p)) = 0.99998 - 0.998p [1 mark]
→ 0.00198 / (0.99998 - 0.998p) = 0.5 [1 mark]
→ p = 0.9980 [1 mark]

15.

a. Let the distribution be X ~ B(n, p):
Mean: np = 4.245 [1 mark]
Standard deviation: np(1 - p) = 1.745^2 [1 mark]
→ 4.245(1 - p) = 3.045025 [1 mark]
→ p = 0.2827 [1 mark]
→ n = 4.245 / 0.2827 = 15 [1 mark]
→ P(2 ≤ X ≤ 6) = 0.8518 [1 mark]

b. For a binomial variable, $P(Y = r) = {}^nC_r * p^r * (1 - p)^{n - r}$;

$P(Y = 2) = P(Y = 3) \rightarrow {}^nC_2 * p^2 * (1 - p)^{n - 2} = {}^nC_3 * p^3 * (1 - p)^{n - 3}$ [1 mark]

Divide both sides by $p^2 * (1 - p)^{n - 3}$: ${}^nC_2 * (1 - p) = {}^nC_3 * p$ [1 mark]

Using the definition of the binomial coefficients in terms of factorials,

$$\frac{n!}{2!(n-2)!} \cdot (1-p) = \frac{n!}{3!(n-3)!} \cdot p$$ [1 mark]

$$\frac{n(n-1)}{2} \cdot (1-p) = \frac{n(n-1)(n-2)}{6} \cdot p$$ [1 mark]

$$\frac{1-p}{2} = \frac{(n-2)p}{6}$$

$$n = \frac{3-3p}{p} + 2$$ [1 mark]

Then, mean = np, so

$$\text{Mean} = np = \left(\frac{3-3p}{p} + 2\right) p = 3 - 3p + 2p = 3 - p.$$ [1 mark]

16.

a. $\bar{x} = \sum x_n / n = 1350 / 60 = 22.5$ [1 mark]

$$\sigma^2 = \frac{\sum x^2}{n} - \bar{x}^2$$ $= (30685 / 60) - 22.5^2 = 5.1666\ldots = 5.17$ [1 mark]

(Also accept 5.25 using sample variance formula)

b. **Define the random variables:**

X = time taken by soldiers in the morning

Y = time taken by soldiers in the afternoon

Let D be a variable for the difference in the sample means:

$D = \bar{X} - \bar{Y}$ [1 mark]

Hypotheses are:

H_0 : no difference in mean time of soldiers in afternoon than in morning (D = 0)

H_1 : a difference in mean time of soldiers in afternoon to in morning (D ≠ 0)

[1 mark]

Deriving the z-statistic:

Variance of D = variance of X-bar + variance of Y-bar, [1 mark] so

$\sigma_D^2 = \frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}$ = $(5.25^2 / 60) + (5.48^2 / 60) = 0.9599$ [1 mark]

(or $\sigma_D^2 = 0.9460$ if using population variance)

So under H_0, (where the mean of X = mean of Y) the *z*-score of *D* is

$$z = \frac{d - \bar{D}}{\sigma_D} = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{0.9599}} = \frac{(22.5 - 24.1) - 0}{0.9797} = -1.633$$ [1 mark]

(or z = -1.645 if using population variance)

Area = 0.025 since two tail

Critical value = $\Phi^{-1}(0.025) = -1.9600$ [1 mark]

Conclusion:

-1.633 > -1.9600 (within acceptance region) so accept H_0 [1 mark]

There is insufficient evidence to suggest there is a significant difference between the mean times taken by the soldiers to complete the course in the morning to in the afternoon. [1 mark]