MATHEMATICS

Stage 5

Statistics and Probability, Data Analysis

- Bivariate Data
- Box Plots and 5-Number Summary
- Standard Deviation

Statistics and Probability, Probability

- Multi-Stage Events
- Relative Frequency and Venn Diagrams

norWEST CHRISTIAN COLLEGE
STRENGTH OF PURPOSE

Teacher: Kyall Duncan

Exam Equivalent Time: 109.5 minutes *(based on allocation ~1.25 min/mark)*

Worked Solutions

1. C Bivariate Data (014201) Core B HSC

Correlation is positive.

NB. The skew does not directly relate to correlation.

$\Rightarrow A$

---

2. C Box Plots and 5-Number Summary (014192) Core A HSC

$\text{IQR} = \text{Upper Quartile} - \text{Lower Quartile}$

In both box plots, IQR = 3 intervals (against bottom scale)

$\Rightarrow D$

---

3. C Bivariate Data (015107) Core B HSC

Positive correlation means that as one variable increases, the other tends to increase also.

$\Rightarrow A$

---

4. C Multi-Stage Events (029225) Core A HSC

Since the two events are independent:

$$P(\text{W}) = \frac{7}{10}$$

$$P(\text{WW}) = \frac{7}{10} \times \frac{7}{10}$$

$\Rightarrow D$

---

5. B Bivariate Data (262967) Core B

Line passes through $(0, 3.5)$ and $(50, 82)$

$$\text{Gradient} = \frac{y_2 - y_1}{x_2 - x_1} \approx \frac{82 - 3.5}{50 - 0} \approx 1.57$$

Distance is the dependent variable (y) and the y-intercept is approximately 3.5.

$\Rightarrow A$

---

6. B Box Plots and 5-Number Summary (002595) Core A HSC

Since the median (155) is closer to the lower quartile (150) and rang
low (140) than the upper quartile (190) and range high (200), it is positively skewed.

$\Rightarrow B$

Mean mark 47%.

Worked Solutions

7. B+ Bivariate Data (002211) Core B HSC

Increased hours training should reduce the time to complete a race.

$\Rightarrow B$

Mean mark 43%.

---

8. B+ Multi-Stage Events (004350) Core A HSC

Let $P(\text{Lou successful}) = P(L) = 0.7,\ P(\text{not } L) = 0.3$

Let $P(\text{Ali successful}) = P(A) = 0.6,\ P(\text{not } A) = 0.4$

$$\begin{aligned} P(\text{only 1 successful}) &= P(L) \times P(\text{not } A) + P(\text{not } L) \times P(A) \\ &= (0.7 \times 0.4) + (0.3 \times 0.6) \\ &= 0.28 + 0.18 \\ &= 0.46 \end{aligned}$$

$\Rightarrow D$

Mean mark 48%.

---

9. A Box Plots and 5-Number Summary (060352) Core A HSC

In History $\Rightarrow$ $Q_3 = 30$ marks

$\therefore$ Scoring over 30 $= 25\% \times 112 = 28$ students

In Geography $\Rightarrow$ Median $= 30$ marks

$\therefore$ Students completing Geography $= 2 \times 28 = 56$ students

$\Rightarrow C$

---

10. A Multi-Stage Events (010151) Core A HSC

$P(\text{rains}) = \dfrac{8}{31}$ (independent event for each day)

Since each day has same probability:

$$P(R_1R_2) = \frac{8}{31} \times \frac{8}{31}$$

$\Rightarrow D$

Mean mark 16%. Lowest mark of any MC question in 2014!

---

11. C Relative Frequency and Venn Diagrams (251080) Path B

	English	No English
French	5	3
No French	8	4

---

12. C Standard Deviation (262472) Core A

Runs scored dataset: $8, 12, 15, 21, 23, 27$

By calculator (using Statistics mode):

Std Dev $= 6.574\ldots = 6.6$ (1 d.p.)

---

13. C Relative Frequency and Venn Diagrams (250997) Path B

$P(\text{man chosen wears watch})$

$$\begin{aligned} &= \frac{\text{number of men wearing watch}}{\text{total men}} \\ &= \frac{25}{60} \\ &= \frac{5}{12} \end{aligned}$$

---

14. C Standard Deviation (262468) Core A

Exam result dataset: $56, 63, 69, 72, 77, 78, 84$

By calculator (using Statistics mode):

Std Dev $= 8.84\ldots = 8.8$ (1 d.p.)

---

15. C Standard Deviation (262529) Core A

i. Dataset values: $1, 2, 2, 2, 3, 4, 4$

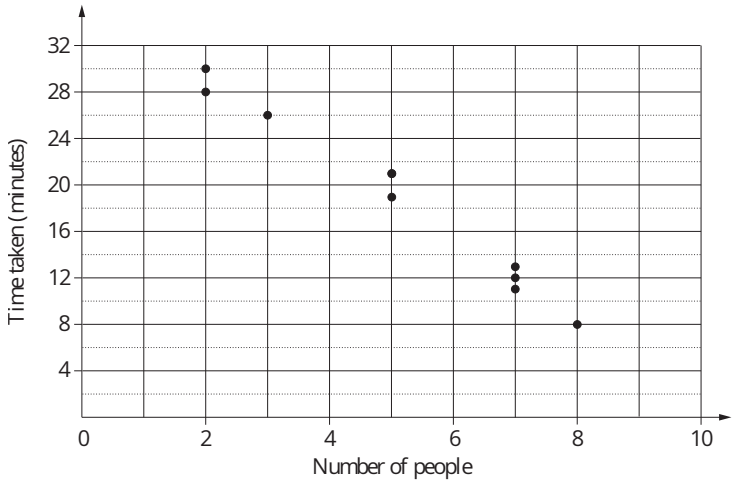
$$\text{Mean} = \frac{1+2+2+2+3+4+4}{7} = \frac{18}{7} = 2.571\ldots = 2.57 \text{ (2 d.p.)}$$

ii. By calculator (using Statistics mode):

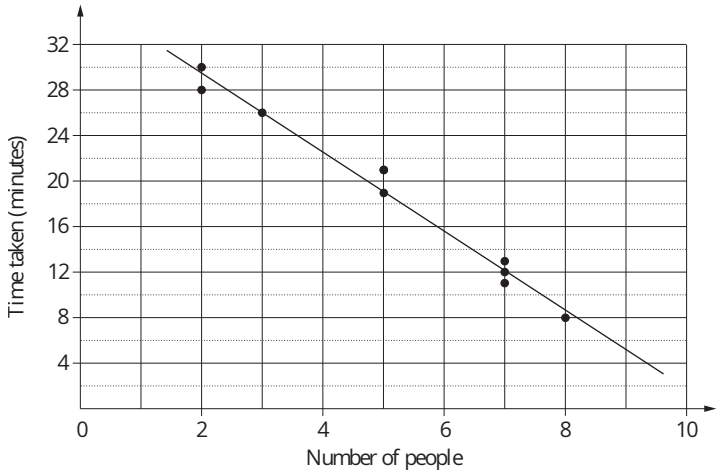
Std Dev $= 1.049\ldots = 1.05$ (2 d.p.)

16. C Bivariate Data (262725) Core B

a.



b.



c. Form: linear (i.e. straight line)
Direction: negative
Strength: strong

---

17. B Box Plots and 5-Number Summary (261221) Core A

i. Ordering the data values:

$3,\ 4,\ 7,\ 8,\ 8,\ 10,\ 12,\ 13,\ 15$

Median $= 8$

ii. $\underbrace{3}_{\text{min}},\ \underbrace{4,\ 7}_{Q1},\ 8,\ \underbrace{8}_{Q2},\ 10,\ \underbrace{12,\ 13}_{Q3},\ \underbrace{15}_{\text{max}}$

Min	Q_1	Q_2	Q_3	Max
3	5.5	8	12.5	15

$\text{IQR} = Q_3 - Q_1 = 12.5 - 5.5 = 7$

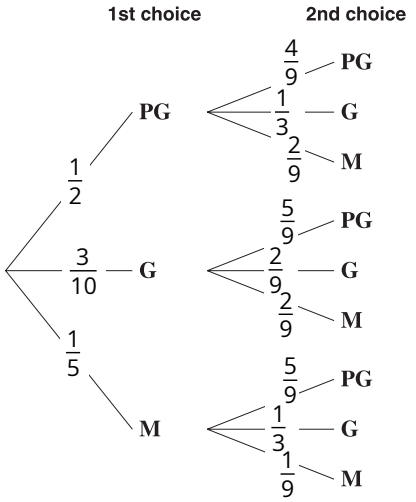
---

18. B Multi-Stage Events (020615) Core A HSC

i. 5 PG, 3 G, 2 M

$$P(\text{M}) = \frac{2}{10} = \frac{1}{5}$$

ii.



iii. $P(\text{same rating})$

$$\begin{aligned}
&= P(\text{PG, PG}) + P(\text{G, G}) + P(\text{M, M}) \\
&= \left(\frac{1}{2} \times \frac{4}{9}\right) + \left(\frac{3}{10} \times \frac{2}{9}\right) + \left(\frac{1}{5} \times \frac{1}{9}\right) \\
&= \frac{2}{9} + \frac{1}{15} + \frac{1}{45} \\
&= \frac{14}{45}
\end{aligned}$$

---

19. B Bivariate Data (263015) Core B

i. 18.5 cm wrist $\Rightarrow$ Ankle circumference ≈ 24.5 cm

ii. y-intercept occurs when ankle circumference $= 0$ cm, which is meaningless in this context.

---

20. B Standard Deviation (262487) Core A

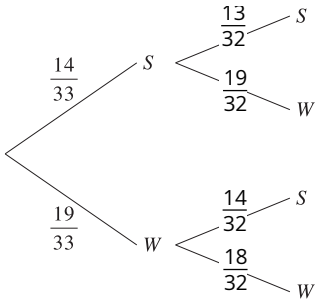
By inspection, Test B has a higher range than Test A.
Test A results look to have a tighter spread of data points about its expected mean, when compared to Test B.
Test B will therefore have a higher standard deviation.

---

21. B+ Multi-Stage Events (003169) Core A HSC

i.

Mean mark (i) 43%.



ii. $P\ (2 \text{ silk}) = P(S_1) \times P(S_2)$

$$= \frac{14}{33} \times \frac{13}{32}$$

$$= \frac{91}{528}$$

iii. $P\ (\text{different}) = P(S_1, W_2) + P(W_1, S_2)$

$$= \left(\frac{14}{33} \times \frac{19}{32}\right) + \left(\frac{19}{33} \times \frac{14}{32}\right)$$

$$= \frac{532}{1056}$$

$$= \frac{133}{264}$$

Mean mark (iii) 41%.
MARKER'S COMMENT: In better responses, students multiplied along the branches and then added these two results together

---

22. B+ Box Plots and 5-Number Summary (005376) Core A HSC

i. Since the median $= 12$ years

$\Rightarrow$ 50% of children are aged 12–18 years

$\therefore$ Children aged 12–18 $= 50\% \times 1750 = 875$

Mean mark (i) 45%

ii. Upper quartile (2010) $= 12$ years

Children in upper quartile $= 875$ (from part (i))

$\therefore$ Children aged 0–18 $= 4 \times 875 = 3500$

Mean mark (ii) 25%

iii. Changes in distribution (include 2 of the following):

- the lower quartile age is lower in 2010
- the median is lower in 2010
- the upper quartile age is lower in 2010
- the interquartile range is greater in 2010
- 2010 is positively skewed while 2000 is negatively

Mean mark (iii) 35%
MARKER'S COMMENT: A number of students incorrectly identified "positive" skew as "negative" skew here.

iv. Implication for government planning:

Since the children are getting younger in 2010,

- Approve and build more childcare facilities
- Build more school and public playgrounds

Mean mark (iv) 46%
MARKER'S COMMENT: Answers should reflect the 1 mark allocation.

---

23. B+ Multi-Stage Events (250900) Core A

$$\text{Percentage} = \frac{\text{number of 9's}}{\text{number of spins}} \times 100$$

$$= \frac{5}{50} \times 100$$

$$= 10\%$$

---

24. B+ Standard Deviation (262338) Core A

i. By calculator (in Statistics mode):

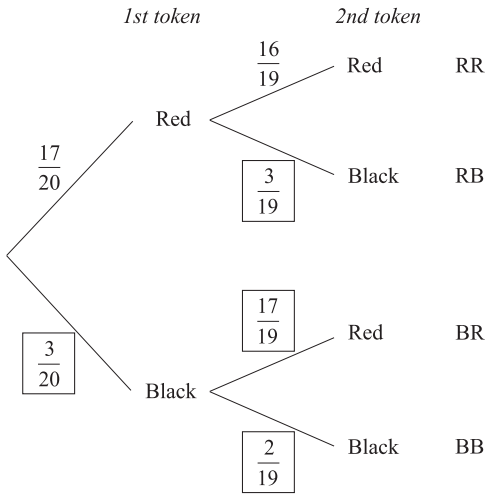
Std Dev $= 1.5947\ldots = 1.59$ (to 2 d.p.)

ii. Standard deviation is a measure of how much members of a data group differ from the mean value of the group.

---

25. B+ Multi-Stage Events (250979) Path B

i.



ii. $P(\text{at least one red})$

$$= 1 - P(BB)$$
$$= 1 - \frac{3}{20} \times \frac{2}{19}$$
$$= \frac{187}{190}$$

---

26. A Box Plots and 5-Number Summary (261404) Core A

i. $\text{IQR}_{(\text{English})} = 80 - 50 = 30$

ii. Skewness

- English has greater negative skew
- Maths is more normally distributed

Mean mark (ii) 35%
COMMENT: The best answers use the correct language of location and spread such as ***mean, median, interquartile range, standard deviation and skewness***.

Location and Spread

- English has a range of 85, Maths has 40.
- English has larger IQR than Maths (30 vs 15)
- Maths' median (75) is higher than English (70)
- Same upper quartile marks (80)
- English has highest and lowest individual mark

---

27. A Multi-Stage Events (004531) Path B HSC

$$P(\text{Mary wins}) = \frac{2}{100}$$
$$= \frac{1}{50}$$

Mean mark 31%.
MARKER'S COMMENT: Very few students calculated Jane's chance of winning correctly. Note the use of "at least" in the question. Finding $1 - P(\text{complement})$ is the best strategy here.

$$P(\text{Jane wins at least 1}) = 1 - P(\text{loses both})$$
$$= 1 - \frac{99}{100} \times \frac{99}{100}$$
$$= 1 - \frac{9801}{10\ 000}$$
$$= \frac{199}{10\ 000}$$

Since $\frac{1}{50} > \frac{199}{10\ 000}$

$\Rightarrow$ Mary has a better chance of winning.

---

28. A Bivariate Data (004584) Core B HSC

i. The correlation between height and mass is positive and strong.

Mean mark 48%.

ii. Using $P_1(40, 1.2)$ and $P_2(80, 10.4)$

$$\text{Gradient} = \frac{y_2 - y_1}{x_2 - x_1}$$
$$= \frac{10.4 - 1.2}{80 - 40}$$
$$= \frac{9.2}{40}$$
$$= 0.23$$

Mean mark 18%.
MARKER'S COMMENT: Many students had difficulty due to the fact the horizontal axis started at $H = 40\text{cm}$ and not the origin.

Line passes through $P_1(40, 1.2)$

Using $y - y_1 = m(x - x_1)$

$$y - 1.2 = 0.23(x - 40)$$
$$y - 1.2 = 0.23x - 9.2$$
$$y = 0.23x - 8$$

$\therefore$ Equation of the line is $M = 0.23H - 8$

---

29. A Standard Deviation (262573) Core A

Standard deviation is a measure of how much members of a data group differ from the mean value of the group.

It follows that the relative standard deviation between two datasets is not affected if $\bar{x}_1 < \bar{x}_2$.

Therefore, the statement is incorrect.

---

30. A Standard Deviation (129518) Core A

i. 30 data points

Mean mark 50%.

$$\text{Median} = \frac{15\text{th} + 16\text{th}}{2} = \frac{4+8}{2} = 6$$

ii. $\text{Lower limit} = 5.4 - 4.22 = 1.18$

$\text{Upper limit} = 5.4 + 4.22 = 9.62$

Mean mark 34%.

$$\begin{aligned} \%\ \text{between} &= \frac{13}{30} \times 100 \\ &= 43.33\ldots\% \\ &= 43\%\ \ (\text{nearest } \%) \end{aligned}$$

---

31. A Bivariate Data (263106) Core B

i. 1st test mark $= 34 \Rightarrow$ 2nd test mark $= 46$

ii. 5 values are within 1 grid height (measured vertically), or 2 marks, from the LOBF.